A STUDY ON MECHANICAL TRANSLATION FROM ENGLISH INTO JAPANESE(Dissertation_全文)

AUTHOR(S):

Sugita, Shigeharu

A STUDY ON MECHANICAL TRANSLATION FROM ENGLISH INTO JAPANESE

by

SHIGEHARU SUGITA

DOCTORAL THESIS, KYOTO UNIVERSITY

JANUARY, 1968

A STUDY ON MECHANICAL TRANSLATION FROM ENGLISH INTO JAPANESE

by

SHIGEHARU SUGITA

DOCTORAL THESIS, KYOTO UNIVERSITY

JANUARY, 1968

PREFACE

The communication between man and machine has become a very important problem to be realized, as the use of computer flourishes rapidly in these days. This communication is desirable to be performed by human languages so that everyone can give instructions to the computer. But the mechanical processing of natural languages is very difficult because of their diversity or flexibility in syntactic and semantic structure. Especially, mechanical translation of natural languages belonging to the different language-families is one of the most difficult problems. This topic, however, is a very interesting computer application to nonarithmetic problems, and a mechanical processing is useful to study many linguistic problems which are

---

ｼﾞｮﾌﾞﾝ

ｺﾝﾋﾟｭﾀｱ ﾉ ｿﾉ ﾂｶｳ ｺﾄ ｶﾞ ｺﾚﾗ ﾋﾄ ﾉ ﾅｶﾆ ｼﾞﾝｿｸﾆ ﾊﾝｼｮｸｽﾙ ﾆ ﾂﾚﾃ ﾋﾄ ｿｼﾃ ｷｶｲ ﾉｱｲﾀﾞ ﾆ(ﾉ) ｿﾉ ﾂｳｼﾝ ﾜ ｼﾞｯｹﾞﾝ ｾﾗﾚﾙ ﾀﾒﾉ 1ﾉ ﾀｲﾍﾝ ｼﾞﾕｳﾖｳﾅ ﾓﾝﾀﾞｲ ﾆ ﾅﾘﾏｼﾀ。 ﾀﾞﾚﾓ ｶﾞ ｿﾉ ｺﾝﾋﾟｭﾀｱ ｴ ﾉ ｷｮｳｼﾞﾕ ﾖ ｱﾀｴﾙ ｺﾄｶﾞﾃﾞｷﾙ ﾖｳﾆ ， ｺﾚ(ｺﾉ) ﾂｳｼﾝ ﾜ ﾆﾝｹﾞﾝ ﾉ ｹﾞﾝｺﾞ ﾆ ﾖｯﾃ ﾂｲｺｳｾﾗﾚﾙ ﾀﾒﾆ ｺﾄｶﾞ ﾉｿﾏｼｲ。 ｼｶｼ ｼｾﾞﾝﾉ ｹﾞﾝｺﾞ ﾉ ｿﾉ ｷｶｲﾉ ｼｮﾘ ﾜ ｺｳﾌﾞﾝﾉ ｿｼﾃ ｲﾐﾉ ｺｳｿﾞｳ ﾉ ﾅｶ ﾆ(ﾉ) ｶﾚﾗ ﾉ ｿｳｲ ｱﾙｲﾜ ｼﾅﾄｶｻ ﾉ ﾀﾒ ﾀｲﾍﾝ ｺﾝﾅﾝ ﾃﾞ ｱﾙ。 ﾄｸﾆ ， ｿﾉ ｺﾄﾅｯﾀ ｹﾞﾝｺﾞｰｶｿﾞｸ ｴ ｿﾞｸｼﾂﾂｱﾙ ｼｾﾞﾝﾉ ｹﾞﾝｺﾞ ﾉ ｷｶｲﾉ ﾎﾝﾄｸ ﾜ ｿﾉ ｻｲﾀﾞｲﾉ ｺﾝﾅﾝﾅ ﾓﾝﾀﾞｲ ﾉ ﾋﾄﾂ ﾃﾞ ｱﾙ。 ｺﾚ(ｺﾉ) ﾜﾀﾞｲ ﾜ ｼｶｼﾅｶﾞﾗ NONARITHMETIC ﾓﾝﾀﾞｲ ｴ ﾉ 1ﾉ ﾀｲﾍﾝ ｵﾓｼﾛｲ ｺﾝﾋﾟｭﾀｱ ｵｵﾖｳ ﾃﾞ ｱﾙ ， ｿｼﾃ 1ﾉ ｷｶｲﾉ ｼｮﾘ ﾜ ﾃﾞｴﾀ ﾉ 1ﾉ ｵｵｷｲ ｶｽﾞ ﾉﾅｶﾆ ｶｸｻﾚﾀ

hidden in a great number of data. This problem may not be attained completely in a short time, but the increase in number of documents or literatures needs rapid processing of them. And now, the mechanical processing of natural languages is in urgent demand of the time.

This doctoral thesis describes the procedure of mechanical translation which converts the syntactic structure of English into the syntactic structure of Japanese. It may not be said to be a real language translation, if semantic information is not taken into consideration. It is difficult, however, to separate clearly semantic aspect from syntactic aspect. This paper does not aim at a completely automatic high quality translation, but intends only machine-aided translation, for the deep structure of meanings is not known explicitely even to human beings. Even if it is restricted to the syntactic translation, the problem is not so easy in case of English-Japanese translation.

In this paper the syntactic structures of English and Japanese

---

オオクノ LINGUISTIC モンダイ ヲ ヘンキヨウスル タメニ ユウヨウ デ アル。 コレ(コノ) モンダイ ワ 1ノ ミジカイ ジカン ノ ナカニ カンセンニ タッセイセラレヌ カモ シレヌ。シカシ ショルイ アルイワ ブンカ(ブンケン) ノ カズ ノ ナカニ(ノ) 1ノ ゾウカ ワ カレラ ノ ハヤイ ショリ ヲ ヒツヨウトスル。 ソシテ イマ シゼン ノ ゲンゴ ノ ソノ キカイノ ショリ ワ ソノ ジカン(ジダイ) ノ ナカニ(ノ) キンキユウ ノ ヨウキユウ デ アル。

コレ(コノ) DOCTORAL THESIS ワ ニッポン(ゴ) ノ ソノ コウブンノ コウゾウ ノ ナカエ イギリス(ゴ) ノ ソノ コウブンノ コウゾウ ヲ ヘンカンスル キカイノ ホンヤク ノ ソノ テツズキ ヲ キジュツスル。 イミノ ジョウホウ ガ コウリヨ ノ ナカエ トラレヌ ナラ ソレ ワ 1ノ ジッザイノ ゲンゴ ホンヤク デ アル タメニ(トワ) イワレヌ カモシレヌ。 コウブンノ メン カラ ハッキリト イミノ メン ヲ ワケル コトワ シカシナガラ、 ソレワ コンナン デ アル。 イミノ ソノ フカイ コウゾウ ガ ニンゲン ノ BEINGS I EXPLICITELY サエ シラレヌ ノデ コレ(コノ) カミ ワ 1ノ カンセンニ ジドウテキノ タカイ ヒンシツ ホンヤク ノ トコロ ニ(ヲ) ネラワヌ、シカシ ユイツ キカイー テダスケセラレタ ホンヤク ヲ イトスル。 ソレ ワ コウブン ノ ホンヤク エ セイゲンセラレル ケレド、ソノ モンダイ ワ イギリス(ゴ) ノ ソシテ ニッポン(ゴ) ノ ホンヤク ノ バアイ(ハコ) ノ ナカニ ソウ ヤサシクナイ。

コレ(コノ) カミ ノ ナカニ イギリス(ゴ) ノ ソシテ ニッポン(ゴ) ノ ソノ コウブン ノ コウゾウ ワ クーコウゾウ

are supposed to be phrase-structure. Of course, there are many structures which can not be treated by phrase-structure grammar. But several experiments and the consideration of verbal behaviour of human beings will show that a large part of the sentences (in scientific papers) have such structure.

One of the characteristic features of the translation method described in this paper is that this method can treat both simple sentences and complex sentences without any distinction. The computer program is separated from grammar so that the main procedure of program is to look the list of grammatical rules. This becomes possible because of the hierarchy in phrase structure grammar which is introduced in this paper. By this hierarchy the context-free style rules can play the same role as context-sensitive rules play. The hierarchy in phrase-structure rules, however, depends on the analysis method (parsing-direction), and also depends on the classification of words and phrases

---

デ アル タメニ カテイサレル。　モチロン、クーコウゾウ GRAMMAR ニ ヨッテ タイグウセラレヌ コトカデキル オオクノ コウゾウ カ アル。　シカシ ニンゲン ノ BEINGS ノ コトバ ノ BEHAVIOUR ノ イクツカノ ジッケン ソシテ ソノ コウリョ ワ (カカクテキナ カミ ノ ナカ) ニ(ノ) ソノ ブンショウ ノ 1ノ オオキイ ブブン カ ソノヨウナ コウゾウ ヲ モツコトヲ シメス タロウ。

コレ(コノ) カミ ノ ナカニ キジュツセラレタ ソノ ホントク ホウホウ ノ ソノ トクチョウ ヨウボウ ワ コレ(コノ) ホウホウ カ イカナル クベツ ナシ ニ(ノ) リョウホウ カンタンナ ブンショウ ソシテ フクザツナ ブンショウ ヲ タイグウスル コトカデキル コトデ アル。　プログラム ノ ソノ シュヨウナ テツズキ カ ブンポウ ノ キソク ノ ソノ ヒョウ ヲ ミル ヨテエデ アル ヨウニ、ソノ コンピュタア プログラム ワ GRAMMAR カラ ワケラレル。

コレ(コノ) ワ コレ(コノ) カミ ノ ナカニ ショウカイセラレル ク コウゾウ GRAMMAR ノ ナカニ(ノ) ソノ HIERARCHY ノ タメ カノウ ニ ナル。　コレ(コノ) HIERARCHY ニ ヨッテ ブンミャクーカンジャスイ キソク カ スル ニツレテ、ソノ ブンミャクージュウナ カタ キソク ワ ソノ オナジ ヤクワリ ヲ スル コトカデキル。

クーコウゾウ キソク ノ ナカノ(ニ) ソノ HIERARCHY ワ シカシナガラ ソノ カイセキ ノ ホウホウ(PARSING ツツアルーホウコウ) ニ タヨル。　ソシテ タンゴ ソシテ ク(CLUSTER) ノ ソノ ブンルイ ニ タヨル。

(clusters). The ordinary analysis of English syntax is begun at the top of the sentence, but the method described in this paper begins at the end of the sentence, because English has right-recursive structure, and Japanese has left-recursive structure. As for the classification of part of speech, it must be performed by considering the correspondence between English and Japanese syntactic structure. Therefore, this method is specialized to English-Japanese translation, but if the number of steps in procedure are taken into consideration, this method is more effective than other universal methods which use common intermediate languages between source-language and target-language.

An experiment, which was carried out with : 400 idioms, 8000 English words, 1000 phrase-structure-rules, and 1300 sample sentences, may not be enough to draw a hasty conclusion about the possibility of English-Japanese mechanical translation even from the syntactic point of view. But it may be said that the hypothesis of phrase structure is

---

イギリス ノ コウブン ノ ソノ フツウノ カイセキ ワ ソノ ブンショウ ノ ソノ チョウシ ヨ ノ トコロ ニ ハジメラレル。シカシ イギリス ワ 1ノ ミギ ノーRECURSIVE コウゾウ ヲ モツ ソシテ ニッポン ワ 1ノ サラレターRECURSIVE コウゾウ ヲ モツ ノデ コレ(コノ) カミ ノ ナカ ニ キジュツセラレタ ソノ ホウホウ ワ ソノ ブンショウ ノ ソノ マツタン ノ トコロ ニ ハジメル。 エンセツ ノ ブブン ノ ソノ ブンルイ ニ カンシテ、ソレ ワ イギリス ノ ソシテ ニッポン ノ コウブン ノ コウゾウ ノ アイダ ニ(ノ) ソノ ソウトウ ヲ コウリョスル コト ニ ヨッテ ツイコウセラレ ネバナラヌ。 ソレユエ、コレ(コノ) ホウホウ ワ イギリス ノ ーニッポン ノ ホンヤク エ トクシュカセラレル。シカシ テツズキ ノ ナカ ニ(ノ) ステップ ノ ソノ カズ ガ コウリョ ノ ナカ エ トラレル ナラ コレ(コノ) ホウホウ ワ ミナモトーゲンゴ ソシテ TARGETーゲンゴ ノ アイダ ニ(ノ) キョウツウノ チュウカンノ ゲンゴ ヲ ツカウ ホカノ ウチュウノ ホウホウ ヨリ オオク コウカテキ デ アル。

400 カンヨウゴ、8000 イギリス ノ タンゴ、1000 クーコウゾウーキソク、ソシテ 1300 ミホン ブンショウ ト(ヲ モツタ) オコナワレマシタ 1ノ ジッケン ワ ケシキ ノ ソノ コウブン テン カラ サエ イギリス ノーニッポン ノ キカイ ホンヤク ノ ソノ カノウセイ ニ ツイテ 1ノ イソギ ノ ケツロン ヲ ヒク タメニ ジュウブン デ アラヌ カモシレヌ。 シカシ ソレ ワ ク コウゾウ ノ ソノ カセツ ガ イギリス ノ ソシテ ニッポン ノ

adequate for English and Japanese syntax, and this method is useful for syntactic analysis and synthesis in natural language processing, if only the determination of working function of words in the sentence is performed correctly.

One of the most important problems which must be studied yet is how to introduce semantic information in connection with syntactic information. Someone may say, however, that it is an "alchemy" in modern times to try to treat "semantics" by the digital computer which is thought to be a monster of abacus.

The Japanese sentences which are written in KANA in the lower part of these pages are the results of mechanical translation of "preface" obtained by the method described in this doctoral thesis. Several corrections (post-edition) are made to inadequate words and expressions.

コウブン ニ タイシテモ ユウブン デ アル コト ヲ イワレル カモシレヌ. ソシテ ユイイツ ソノ ブンショウ ノ ナカ ニ(ノ) タンゴ ノ ハタラキツツアル キノウ ノ ソノ ケッシン(ケツテイ) ガ オコナワレル トラ コレ(コノ) ホウホウ ワ シゼン ノ ゲンゴ ショリ ノ ナカ ニ(ノ) コウゾウ カイセキ ソシテ ソウゴウ ニ タイシテ ユウヨウ デ アル.

マタ ヘンキョウ セラレ ネバナラヌ ソノ サイダイノ ジユウナ モンダイ ノ ヒトツ ワ コウブン ノ ジヨウホウ ト(ヲ モツタ)(トノ) カンケイ ノ ナカ ニ(ノ) イミノ ジヨウホウ ヲ ショウカイスル ホウホウ デ アル. ダレカ ワ ソレ(ソレヲ) ABACUS ノ 1ノ カイブツ デ アル タメニ(ト) オモワレル ソノ ジジタル コンピユタア ニ ヨツテ (イミロン) ヲ タイグウスル コトヲ ココロミル タメノ(コトヲ) ゲンダイノ ジカン(ジダイ) ノ ナカ ニ(ノ) 1ノ (ALCHEMY) デ アル コトヲ シカシナガラ イウ カモシレヌ.

カク ペエジ ノ ソノ ヒクイ ブブン ノ ナカ ニ(ノ) KANA ノ ナカ ニ カカレル ソノ ニツポン ノ ブンショウ ワ コレ(コノ) DOCTORAL THESIS ノ ナカ ニ(ノ) キジユツセラレタ ソノ ホウホウ ニ ヨツテ テニイレマシタ (ジヨブン) ノ キカイ ノ ホンヤク ノ ソノ ケッカ デ アル. イクツカノ テイセイ(ユウヒンーノチ EDITION) ワ フテキトウナ タンゴ ソシテ ヒョウゲン エ ツクラレル.

CONTENTS

———— . ———— . ———— . ———— . ————

Chapter 1

INTRODUCTION

"In the beginning was the Word", says John at the beginning of Gospel. The "Word" in it, however, may not mean human being's languages which are uttered in vocal sound or written using several kinds of alphabet, but it metaphorically represents the state of nature or its image reflected in our brain. So he says "the Word was God".

If it is possible to suppose that there exists harmony and regularity in the "Word", that is, in God or Nature, it is also possible to believe that there exists harmony and regularity in our human languages. Because human language originally grew to notify the existence of some objects in the natural world or to inform other people of the structure or relation in the natural world, and images in one's brain. This is the fundamental nature of human communication.

Human language, however, would not show its graceful shape before us, though it had not less harmony and regularity than the language of mathematics which transcribes the structure of natural world very elegantly. Since olden times, many philosophers and linguists have been trying to strip the veil from human language, but to their disappointment and, at the same time, to their great surprise, they have taken off a surface veil only to find a new veil had been ready under the old one. This characteristic, that is, the inability of complete grasp of the natural shape, may be an essential feature of human language, and is very misterious to us human beings, for it was man himself who constructed language.

Modern science and technology, however, are going to strip all

veils by force without any mercy, using the electronic computer which even God did not make. One of the approaches to this purpose is the problem of "mechanical translation"(MT) of natural language, which keeps step with the progress of digital computer and has a history of two decades. To translate one language into another is very difficult even for human beings, who have been learning languages these fifty thousand years, therefore it may be thought to be impossible for a machine. But the recent progress in linguistics, physiology, psychology, and also electronic engineering, gives us a faint hope of the actualization of mechanical translation. It is true that in mechanical translation it is the machine which translates, but a machine by itself can not do anything, it working only when we give definite procedures to it. In this sense, the mechanical translation depends on the linguistic knowledge of human beings. At present, however, we can not say that mechanical translation is realizable, though, of course, there is no reason why it is impossible. This is only because we can not yet discover the harmony and regularity in languages, and not because the present computer does not work well enough.

Nevertheless, if we think of the fact that many people who studied mechanical translation early in its history gave up the idea of "full automatic high quality translation", it is necessary to reconsider the simple belief that there exists harmony and regularity in any human languages. The full automatic translation may be impossible, or it may not be of practical use, if possible, because of the cost or time needed. Even if it is true, we must make a through study of the structure of natural languages to find the reason why it is impossible or not practical. There must be many hidden aspects of languages which are not yet investigated by human beings because of their complex

structures.

The problem the author raises here is whether it is possible to transform the syntactic structure of one language into that of another one only using syntactic information. The answer to this is negative without any trials, but to know the degree of impossibility is very interesting, and the author intends to investigate what kind of information other than syntax is effective or in what way we can make good use of syntactic information. Therefore, here, a practical use of mechanical translation is not necessarily the sole object of this study.

The author believes that human languages reflect the mechanism of human thinking in the bottom of the structure. The mechanism of thinking is perhaps common to all people of different nationalities. If we stand on that ground, there must lie some common features at the basis of several languages which apparently look different. From the metaphysical point of view, it is natural to think like above. There are, however, some different points on the surface of human thinking and natural language. What is it, then ? Persuing this different points and looking for the deep structure, we will be able to know little by little the real structure. Up to this time many philosophers, linguists, psychologists, phisiologists, and also ethnologists have studied various languages from many angles, and make several experiments about them, so that the hidden part of languages have partially been brought to light. And now, with the aid of computers we can make new steps to analyze all parts of languages and to ask what is the essential difference among them.

As the first step, the author takes up the computer analysis of the syntactic aspects of English and Japanese. Concerning the syntax of English or Russian, a lot of papers have been already written by

foreigners in various manners. Indo-European languages have nearly the same syntactic structures, and so even translation needs not so complex algorism. On the other hand, concerning to Japanese which, at first glance, presents very different syntactic structure (and also semantic structure) from Indo-European languages, it is not too much to say that there are no more than ten papers even in Japan, and there are much less papers which present explicitly the grammatical rules by which the syntactic analysis and synthesis of both English and Japanese can be carried out mechanically. In this paper such grammatical rules are listed and explained.

Though English and Japanese have not linear correspondence in word order and one-to-one correspondence in word itself, we can find in a sense that they have very similar syntactic structure when seen from a certain standpoint. That is, when the result of immediate constituent analysis is depicted in tree diagrams, both trees of typical sentences look like the same except for some branches. In this case the branches of a tree do not intersect each other. The language whose branches in structure trees do not intersect is called "phrase structure language", and the treatment of this is easy even for machines. It is very doubtful, however, whether all languages in the world, or at least English and Japanese, have tree structures or not. But it is probably true that in the semantic or concept trees, if they can be expressed in trees, the branches may not intersect, because otherwise it is very difficult for us to conceive the other elements squeeze themselves in the unity, in other words, we recognize the unity as a whole and can not divide it into some blocks and put the different kind of words into them like sandwich. If it is reasonable to think that semantic aspect of language may affect the syntactic structure,

human languages can be supposed to be "phrase structure". It may be such phisiological or psychological characteristics of human recognition mechanism which make it possible to process natural languages mechanically.

In this paper it is supposed that English and Japanese have phrase structure in syntax, and this hypothesis is, the author believes, largely verified by the experiment.

Now, in modern society, mechanical translation, as well as other mechanical processing of natural languages, are coming to be not only an interesting problem, but also a necessary problem. Because it is true that by the aid of electronic computers human culture and civilization can progress increasingly, if only every man can use computers and communicate with them as in ordinary conversation using human languages, though it is not necessarily an incorrect opinion to regard the computer only as a kind of abacus in the sense that it works only according to the instructions provided by human beings. It is indeed true, however, that in arithmetic calculation or bussiness calculation including its tabulation, such programming languages as FORTRAN, ALGOL, COBOL, or PL/1 are rather more convenient than natural languages, and in the processing of language-like data out of a special symbolic system or the simulation of a certain system, it is also convenient to use symbol manipulating artificial languages, such as COMIT, LISP, SNOBOL, ERIZA, or DYNAMO etc. But if we can not express our questions in arithmetic equations or extremely restricted symbol strings, for example, in the case of information retrieval such as document retrieval or fact retrieval or content analysis etc., we can not but ask question in natural languages. In this situation when asked in a natural language, computers must be fully equipped with various kinds of technique or

algorithm to analyze input sentences, retrieve the desired information, and answer in a natural language. There are, however, two kinds of system. In one system, the computer works only as a transformer of symbols, and interpretation of contents is done by man. In the other system, the computer works both as a transformer and as an interpreter. Mechanical translation, though it is the most difficult problem in natural language processing, may be looked on as a transformer because both terminals of the system are connected with men, and the machine only exchanges the input symbols into output symbols without, as it were, knowing what these symbols represent. Of course transformation is not so simple a code inversion as in data transmission, but it requires as complex algorithm as in the case of interpretation, though such content-understanding is somewhat different from that of the natural language programming or information retrieval. Nevertheless, the problem of mechanical translation involves all sorts of techniques contained in the mechanical processing of natural languages, such as construction of word dictionary, its effective use, structure analysis and synthesis technique, or editing etc. Then, to study the problem of mechanical translation plays a central role in the mechanical language processing systems.

Nowadays people are interested to see what kind of jobs the electronic computer can not do. They say that computers are inferior to men in language processing, game-playing, pattern recognition, and creative power including art-activity. The evaluation of these subjects except creative power depends on the time required and the quality of the results. If there is not time-limitation, even mechanical translation may not be so difficult in both syntax and semantics. Of course, in semantic aspects, the possibilities of interpretation become as many as nearly infinite, but ordinary usages do not vary in so many

cases. If the examples of possible usage are supposed to be stored, though it may come to an enormous amount of information, the machine will be able to research needed information, if given sufficient time. In such a system the mechanization of data-input plays a central role. But this system takes too much time to be practical and of interest. Therefore, time-reducing technique is the main subject in computer application to non-arithmetic problems.

English-Japanese translation methods which were presented up to now have been experimented by using a rather small number of words or rules, and they treated only simple sentences, therefore, even if they got good points for such samples, it does not guarantee their effectiveness. In this paper, the algorithm is tested in almost all real situations, using complete sentences ramdomly selected from several fields, and word dictionary containing 8000 English words.

Chapter 2 describes a short history of mechanical translation and several problems which prevent this subjects. In chapter 3, somewhat ideological contemplations on phrase structure in natural languages are given, though they are rather incomplete. Also in chapter 3, the criteria for classification of parts of speech and priority of rewriting rules are described, and a list of all rewriting rules and explanationtion of some of them are given. Chapter 4 describes the structure of each dictionary, and some of examples are given in lists. Chapter 5 is concerned with the algorithm of mechanical translation, and a few concrete examples which were printed out by the computer itself are given. In chapter 6 the results of experiments are analyzed.

Chapter 2

HISTORY AND PROBLEMS IN MECHANICAL TRANSLATION

2.1 History and necessity of mechanical translation

It is not so queer that the idea of mechanical translation was brought up by "decipher" in the World War II, for the object of decipher is to translate from an unknown sentence into the known one. The modern technique of mechanical translation, however, is not similar to that of decipher at all, which transforms some sequences of letters of the alphabets or numerals into sequences of meaningful words by aid of a statistical knowledge and a decision on the circumstances where the code was obtained.

In the early history of mechanical translation, say in 1946, A.D. Booth and W. Wiever discussed, in the analogy of decipher, the applicability of a digital computer to language translation. In it the meaning of the word "translation" was only to substitute words in one language for the corresponding words of another language, that is, word-for-word translation. Between Indo-European languages, even this word-for-word translation might be useful (though not enough), but it was noticed that syntactic analysis is necessary even to substitute word for word, and mechanical translation came to be studied from the fundamental linguistic points of view, departing from the analogy of decipher.

There were a few demonstrative experiments of mechanical translation on rather a small scale between English and Russian in the United States of America and also in Russia, in 1954 and 1955; in those days word-for-word translation were yet the leading idea, and the study was directed towards the construction of a mechanical dictionary: classifi-

cation and processing of conjugations, reduction and compression of information in order for effective use of computer memory. The method of syntactic analysis was considered for only one language so that it was not a translation oriented method. If the two languages belong to the similar language family, these monolingual analysis might even be sufficient, but when we treat different languages from different families, the relation between the analysis of the object language and the synthesis of the target language must be considered. That is, the analysis method of the object language will depend on the structure of the target language.

It was not until N. Chomsky published the idea of new type approach to linguistics, that is, phrase structure grammar, that the fairly systematic treatment of natural language became possible. It was in 1957. The emergency of phrase structure grammar brought not so much the new analysis method as the new descriptive tool which consolidates the traditional analysis method from a cognitive point of view.

But almost all methods presented so far, which treat natural languages (mainly Indo-European languages), stand on the hypothesis that their object languages have a phrase structure grammar. Several famous analysis methods, such as Kuno-Oettinger's "Predictive Analysis", Hays' "Dependency grammar", Bar-Hillel's "Categorial grammar", and Chomsky's "Immediate Constituent analysis", etc., are verified to be equivalent each other in the sense that they can be accepted by "the push down store automaton" which was introduced by N. Chomsky. The method presented by P. Phodes which is based on the lattice theory belongs to the dependency-analysis method, and is no exception to the phrase structure grammar. Any languages, however, can be looked on as a phrase structure, if sufficiently large units of concept are taken

for units of grammar. But we are not interested in such a treatment. Our interest exists in the fact that small units (which correspond to "word") can construct sentences by using comparatively few numbers of rule. Phrase structure grammar, however, can not analyze all sentences, if its units are restricted to elementary concept. Then a transformation grammar is introduced to support phrase structure grammar. But the transformation grammar is, as it were, a list of irregular expression, and seems to be not so intrinsic rules in natural languages, since there is no necessary reason why such transformation rules work well, though the rules are useful from the practical point of view.

In the course of the history, there were presented some ideas of making an intermediate language between object and target languages to facilitate easy and effective translation. There are, however, probably no such intermediate languages which human beings can understand and also the machine can easily treat. Such a system as symbolic logic may partially be of use, but among the natural languages there are no candidates. Even if an artificial language which is similar to a natural one, such as Esperanto, is constructed, it may not be used readily as a world language because such a language can not largely depart from the existing languages.

Though to create quite a new artificial language may be impossible, there is a comparatively practical solution. That is, since natural languages have phrase-structure or immediate-constituent structure in most (not but whole) parts of them, then the sentences which are to be mechanically translated or processed have only to be transformed into wholly phrase-structure from by human editing before mechanical processing, or to be written in phrase-structure at the beginning. Those edited languages may not be called intermediate language, but

recently pre-editing and post-editing are coming to be a leading idea in place of an intermediate artificial language. As for the concrete examples of pre- and post-edition, however, few papers describe them. In this paper some examples of edition are shown in later sections.

Linguistic problems have three aspects: phonology, syntax, and semantics. In the so-called mechanical translation, phonology is only treated in connection with word inflection, and not so much studied. Semantics is the most important and involves so many difficult aspects in human communication, and without intensive study of it mechanical translation will not be realized. But our concept of meanings is too vague and too complex to grasp the underlying essential relation, to say nothing of processing by the computer. It is only since about 1961 that the study on semantics especially in relation to mechanical processing are actively investigated, but it is apt to diverse rather to general semantics than to go hand in hand with syntax and mechanical procedure, so that the problems become more and more difficult, and there are no promissing systematic treatment of meanings at present. In this paper semantic problems are not explicitly considered, but in connection with syntax several points are discussed in some sections.

Generally speaking, the present state of mechanical translation can not be said to be progressive. As for Russian-English translation, it is almost near practical use with the aid of human pre- and post-edition, but editing is too tedious and takes much cost and time even in such translation between somewhat similar languages.

As regards English-Japanese translation, the first paper appeared in 1958 but it only treated simple sentences. Since then, not more than five papers have been published, and they only deal with very limited sentences and are not aiming at practical use, though it is undeniable

that they gave an impetus to the study of computer application to natural language processing. In Japan, to my great regret, mechanical translation is paid attention to only as an interesting application of the computer, and not studied by a large group trying to put it to practical use in near future.

2.2 Mechanical translation and Information Retrieval

Is mechanical translation really needed? It is said that the translation of scientific papers, titles of books, panphlets or abstract papers are in urgent demand. Title translation, however, can be done by use of the word dictionary without such a complex technique as mechanical translation. As for scientific papers, not to say literary works, it seems to be impossible to translate them as well as human beings do, especially between English and Japanese. Then the remaining possibility is to translate abstracts only incompletely. (Completeness can not be hoped for forever). Not that sentences in abstracts have different structures from ordinary sentences and are less difficult, but that abstracts are often used to know whether there is any need to read through the whole-paper or not. Therefore, its rule is to attract the reader's interest, so that ambiguity in syntax and semantics, if any, will be allowed. Then, the criteria for translation are loosened, and even mechanical translation is of use in many cases.

If, however, the aim of abstracts is to select the primary documents, then the information retrieval system will give an answer to this question instead of mechanical translation. That is, if one gives

some words concerning his interest or request without reading abstracts or selecting the original paper himself, the information retrieval system will directly take out the desired documents to him. In this system, the key word index takes the place of abstracts, and plays their part as secondary documents representing also semantic relations in the original one. As seen in the above mentioned case, the completeness of the information retrieval system will eliminate such human activities as reading and understanding abstracts and to select primary documents, and the desired literatures can be obtained directly. And the algorithm of comparing input words with the stored index is mainly to consult the word dictionary or thesaurus, and to analyze simple semantic relations.

As long as these uses of abstracts are concerned, therefore, mechanical translation is not necessarily an indispensable system, and it remains only an interesting application of the computer. But in such a country as America, where scientific papers or official documents in Russian or French must be processed in haste, the computer-aided translation which needs post-editing and is slightly faster than human translation is now in practical use. These machine-aided translations will go with the tide from this time forth.

Although mechanical translation can not stand by itself as a complete system without any human aids, the relation between the computer and human language will become probably closer and closer from now on. The operations of mechanical translation include all sorts of problems concerning natural language processing; therefore, the studies of them are fundamental to the general field of mechanical processing of natural languages. Perhaps the studies contribute also to linguistics, that is, phonology and syntax and semantics. The true value of human languages,

however, may exist where a machine can not intervene, but the computer can investigate a lot of linguistic data which human beings can not process, and discover the latent regularity especially in syntax and phonology. Recently computers are going to be applied to the study of literature, such as author identification, writing tendency of a certain author, statistical analysis of literatures, deciphering of unknown language, and so on.

2.3 Problems in mechanical translation

It is the problem of "ambiguity" that people admit as a reason why mechanical translation is difficult. There are syntactic ambiguities and semantic ambiguities, and the close connection between these two ambiguities makes the problem more and more complex.

As the first ambiguous case which belongs to the syntactic one, a problem of unique determination of parts of speech is considered. The word "part of speech" here does not necessarily mean the real function of a word in the sentence, but only the symbol used as a clue for judging the function of the word. For example, in the sentence "They are flying planes.", the word "flying" may be used as an adjectival word or as a progressive form. They are determined by the context in the sentence, the word "flying" being recognized as the ing-form of the verb "fly". In the above example it is certainly determined by its form. On the other hand, in the sentence "The information research requests....", the word "research" and "requests" can not be definitely determined as nouns by their morphological appearance or extremely

limited syntactic context alone. If these parts of speech are not determined, their functions can not be determined, either, and so the syntactic analysis will not proceed. In such cases where one word has two parts of speech, if that word is situated just before or just behind a special word, the selection of adquate symbols may be possible. For example, in "The request...", the word "request" can be determined as a noun because a word just behind a determiner is looked on as a noun rather than a verb. Prepositions and auxiliary verbs, as well as determiners, give a clue to settle the parts of speech of multifunction words, if only the words appear near such key words, otherwise there is no effective algorithm to choose the correct one out of two possibilities, noun or verb.

One method is to try for all possible combinations, but the number of trials increases exponentially as the number of multi-function words increases; nevertheless they are not necessarily good solutions to these problems. Human beings, however, generally can choose correct ones by looking at the wide range of sequences of parts of speech, without knowing explicitly their meanings in many cases. For example, in the next sentence, the part of speech of "help" can be determined as

```
" Since it requires understanding of the content of the
  C2   N3  VT/N1        GT       P2 DT    N1    P2 DT
 document it cannot be regulated to clerical help. "
   N1    N3   VA   VL    PT     TO    AO    VT/N1
```

N1(noun)*. The criterion for this case may be compound of some features: (1) there exists a word which is clearly a verb, (2) it comes just behind an adjectival word and the adjectival word is preceded by a preposition "to", (3) it appears at the end of the sentence, etc.

---

* Symbols for part of speech are explained in Table 3.4.1.1 in section 3.4.1.

The above example is rather a simple one, but in the next sequence of parts of speech which corresponds to the sentence "the only words in that

DT B1 N1 P1 TH N1 TH VT VL PT VL N1 C1 N1
VT N1

request that need be considerd are malignance and stomach", it is not so easy. Those key words DT(determiner), P1(Preposition 1), TH(that), VL(be-verb) etc. are not useful in selecting one of the two symbols for the underlined words. In this case, man will count the number of verbs and relative pronouns, and investigate the relation between them, or he will generate a sentence putting an appropriate word in each position. This process may differ from case to case, and utilize the semantic information as tacit consent, so that it is very difficult for the machine to perform. Then the question is whether there is an effective method to infer the roles of words not only from determiners or auxiliary verbs but also from several words which are situated before or behind the ambiguous words, avoiding the case-by-case procedure. For this purpose, it is probably effective to use the connection table of parts of speech, which corresponds to digrams or trigrams in letter combinations. That is, by using fairly large samples, we make such a frequency table of symbol connections as shown in Table 6.2.2in section 6.2 and when ambiguous words appear, the most probable case is determined mechanically from the probabilistic point of view by investigating in the table the possibilities of a certain sequence including an ambiguous word. But our human languages may perhaps not obey the statistic rules, and so this method does not give a complete solution, though it is practical in some degree.

Now, even if the parts of speech are determined uniquely for each

word, its role in the given sentence can not be determined. It is the wide-range context which determines its real grammatical role in the sentence; however, there are many ambiguous cases which make the operation very difficult. Here the word "ambiguous" is used in a broad sense; that is, in the case of "They are flying planes", the word "flying" is certainly ambiguous because both adjectival and progressive uses are semanticaly possible, but in the next sentence

" In typing isolated words a columnar layout is equally satisfactory to a typist."

the word "isolated" works only as an adjectival word, but it is difficult to determine by the restricted context whether the word "isolated" has an adjectival use; even in such a case the word "isolated" is said to be ambiguous from the syntactic point of view.

It is the cases of prepositional phrases which appear most frequently in an ambiguous way. For example, the next two sentences have quite the same syntactic structure, but in (1) the prepositional

I bought a car with no wheel. (1)

I bought a car with my money. (2)

phrase (underlined part) is used as an adjectival phrase, and in (2) it is used as an adverbial phrase. Therefore, it is impossible to decide whether it works as adjectival phrase or adverbial phrase, simply by using the syntactic information. The above difference between (1) and (2) comes from the fact that human beings interpret any sentence according to the association with the word meanings. If we let the computer perform this human-like processing, a large amount of detailed information about concepts of words must be given to the

computer. It is, however, almost impossible to give machines all sorts of information considering possible uses of various words. Therefore, it may be advisable to begin the semantic study in connection with such syntactically ambiguous cases.

Furthermore, difficult problems for the machine arise in the case of inserted phrases or clauses. The grammatical roles of such phrases are determined by the semantic relations rather than syntactic ones. For example, in the next sentence, the phrase led by "together" is a

> " New results are presented, offering insight into the performance and optimization of linear and adoptive delta modulation, together with a comparison with pulse code modulation."

syntactically ambiguous case. In the following example, the phrase

> " This report describes the results of a study of statistical delta modulation, a new method of digital transmission of analog information. "

"a new....information" is in apposition to the phrase "the results...", but in the next example which is of almost the same structure in syntax,

> " This book describes very interesting story, Jack. "

the noun phrase "Jack", which corresponds to "a new...", has no relation to "very interesting story". The latter example may not be exactly analogous to the former one, but there are many such cases.

At any rate, syntactic information is not powerful enough to determine how a set of words relate to other words. This freedom of syntactic structure is one reason why the language can express infinitely many affairs using only a finite number of symbols and rules. On the other hand, this fact prevents the machine, which operates in a

deterministic manner, from processing natural languages more effectively.

If, however, it is the characteristics of languages that the structure can not be determined uniquely by the syntax, then it is perhaps possible to translate them by taking it the other way round. That is, in Japanese, all adjectival and adverbial phrases or clauses come before the words to be modified, so even if the roles of phrases differ from sentence to sentence, the word order in both languages is the same, though particles which must be added to the phrases are different according to the roles of the phrases. For example, in the next two sentences, a prepositional phrase in (a) is adjectival, and in (b) it

I saw a flower in the vase. watashi wa kabin no naka no hana o mita. (a)

I put a flower in the vase. watashi wa kabin no naka ni hana o ireta. (b)

is adverbial, and the only difference is their particles NO and NI. Therefore, if the syntax in English can not be clearly analyzed, it is probably a good idea to insert ambiguous particles in Japanese so that it can be interpreted in two ways. In the above case a particle NI(NO) is to be inserted. In other words, it is to translate source language into target language, preserving the ambiguity in the original text as it is. This idea is analogous to code inversion in information transmission. But in complex sentences it becomes difficult even to preserve ambiguity as it is; namely, the translated sentence includes more ambiguity than the original sentence does because of the intervention of noise information; here by "noise" is meant wrong word order exchange and insertion of ambiguous particles. These facts make mechanical translation more difficult. Especially in English-Japanese translation, exchange of word orders brings about outrageous translation, so that in

some cases post-editing becomes impossible. For example, in the next translation, the result is obtained by application of locally correct rules, but it is almost impossible to apply post-edition to this trans-

(English) An ideal natural language information system would permit storage of documents and their retrieval by normal human process, in such a manner that the user could be unaware of the existence of the operative computer system.

(Japanese) sono tukawu koto ga sono OPERATIVE konpyutaa soshiki no sono sonzai no siranai kotoga dekiru sonoyona 1no taido no naka ni 1no risono sizenno gengo jyoho soshi ki wa syorui sosite hutuno ningenno katei ni yotte karerano kensaku no chikuseki o yurusu darowu.

lation without reference to the original text.

It goes without saying that semantic ambiguity is the intrinsic characteristic of the natural languages. Most of the cases belong to the selection problem of multivocal words. This problem is difficult to be solved even in sentence-to-sentence translation. For example, the meaning of the word "paper" in "The paper describes the fact... can not be judged whether it is "newspaper" or "thesis" by this simple sentence only. Also in the example "There are two principal tables in the logic file.", the word "table" means a "list", but not "desk", but there is perhaps no algorithm to select a word which means "list". If there are several multivocal words in a sentence, to select the appropriate translation is very difficult, even if plenty of meaning-information is given to each word. In these cases man must select the correct one.

In semantics, as well as in syntax, the criteria for selection of words and rules must be changed from case to case, and to judge when the process must be changed is a very difficult problem. The main reason why mechanical translation is difficult consists in this flexible character of natural languages. As for a human being, he communicates with each other not only by the system of symbols but also through the real world which the symbols represent: therefore, slight changes in word order or existence of multi-meaning words do not prevent him from finding the correct situations. We must analyze and abstract the relation hidden in the real world to make mechanical translation possible; however, this is tremendously difficult if not impossible.

One of the promising ways to the study of semantic problems is to clarify the mechanism of analogical inference and metaphor, which is the leading process in human recognition or perception activity (association). It is also important to study how the contents or meanings of sentences can be recognized, and how they are expressed in another language by using another method than by the simple one-to-one correspondence for word and syntax.

Chapter 3

PHRASE STRUCTURE AND ENGLISH-JAPANESE TRANSLATION

3.1 Phrase structure in natural languages

The concept of "phrase structure" or "immediate constitutent structure" is the characteristic feature of natural language in syntax which is the most beautiful and plays a leading role in natural language. This suggests that natural languages are generated and developed by human beings and they reflect the physiological or biological aspect of human recognition and perception. Therefore, the study of language gives an important clue to make clear the various human activities.

Phrase structure(PS) and immediate constituent structure(ICS) are the structures in which, as shown in Fig.3.1.1, adjacent elements are

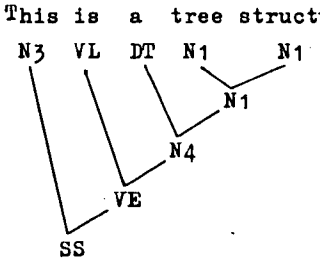


Fig. 3.1.1
Phrase structure.

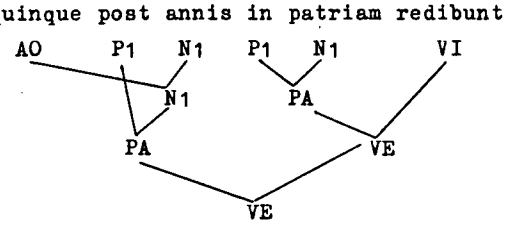


Fig. 3.1.2
Non phrase structure.

connected into larger units, and elements which are apart from each other do not form a group directly. The structure shown in Fig.3.1.2 is not a phrase structure. Such a structure shown in Fig.3.1.1 is called "tree", in which branching points are called "node" and the

tips of the branches are "terminals". The tree structure can be expressed by the set of rewriting rules which have such forms as below,

Fig. 3.1.3 **Rewriting rule and tree structure.**

Where α, β, and γ are node symbols (or non-terminal symbols) or terminal symbol. This rule means that adjacent elements α and β are connected into γ. These sets of rules are equivalent to the tree expressions where no branches intersect each other.

Though ICS and PS treat the same character of languages, ICS attaches great importance to analysis, and PS to synthesis or generation. That is, in PS, the rewriting rule has the inversely directed arrow comparing to that of ICS, and generates a tree using substitution rules.

Fig. 3.1.4 **Generating rule and tree structure.**

The tree generated by PS grammar may cover all trees reduced from natural sentences by ICS grammar; in other words, PS grammar generates too many trees which do not belong to the structures of natural languages.

Now, is it truly adequate to look upon the structure of a language as PS or ICS? The discussion of several points is given below.

Roughly speaking, languages can be thought to be kinds of tools, and in order to communicate with each other in such languages of the present form is not necessarily needed. If one wants to transmit to others the existence of matter or object, he can attain it by bringing the object itself or its equivalent before other people, or by bringing

them to the place where the matter exists. If it is difficult to move the materials, he can express them by pictures. It is shown in the fact that ancient communication was done by means of pictures on the stone or wall, etc. The object itself or stones with pictures drown on them, however, can not be moved freely in time and space, and also they take much space and time to express themselves. Then, gestures, which do not take much space and time and might co-exist with picture painting, came to be used independently. And it is probably true that vocal sound which acompanied gestures as a supplementary tool took the place of gestures. Voice, however, goes down, as it were, one dimensional scale. So it is difficult to retain fixed images continuously, unlike objects or pictures. Then there appeared letters which represent vocal sounds. Letters need two-dimensional space as pictures do, but they consist of combinations of a comparatively small number of symbols which correspond to sound units, so an expression represented by letters can be looked upon as one-dimensional as a flow of thinking. Therefore, the sentence or sequence of letters can be thought to preserve its relation to materials in three-dimensional space or pictures in two-dimensional space. Then its structure can not be quite free. That is, language reflects the nature of human recognition of real world. The ability for recognition which is ascribed to physiological structure is common to all human beings and does not depend on race or age. Then there must be something common to every language which looks different at surface.

First, there is an undividable unit for our recognition. Though this unit differs in size for the object to which we pay attention, disjoint materials or mixture of different kinds of material can not be thought to be a unit, but continous and homogeneous ones are recog-

nized as fundamental units. The word "unit" or "material" is not used to mean chemical substance which consitutes the matter, but it means a functional role. Second, the spacial distance and arrangement, or functional relation of separate units are paid attention to. Though there coexist various kinds of phenomena and subjects in the real world, only some of events or objects in which we have interest rise to the surface of our recognition. This fact is well illustrated in the cocktail party effect in hearing or figure-ground perception in Lebin's experiment. The psychological explanation of these phenomena can be summarised into two fundamental concepts: (1) contiguity, events or matters which are situated near in distance or time are recognized as intimate in relation to each other, (2) analogy, events or matters which are analogous in figures, size, or function are recognized as intimate in relation to each other. The cause-result recognition follows from the contiguity of their times of occurrence. The images of materials or events in space perhaps are stored in our brain in two-dimensional forms rather than in cubic forms. Every thing is remembered by abstracting its characteristic relation from the view point of each one's interest. That is, a certain subject which is paid much interest to is set in the central position in the brain, and other related matters are arranged in plane around it in the positions determined by their relative distance (or psychological distance) from the central one. In this statement the word "position" or "distance" does not mean a physical one, but rather means a mental one. They are connected by the association mechanism. To represent this two-dimensional relation or image by a one-dimensional flow of letters, it is necessary to give a hierarchy to each part of the plane, because in a one-dimensional expression two parts can not be mentioned at the same time: one

of them must be expressed before others. But the hierarchy in the plane figure is not definitely determined, and nor it is linearly ordered. Therefore, the word order in languages is not so intrinsic. If materials or concepts, A and B, are connected with each other in space or functional role, then the representatives of A and B must be situated also in intimate positions and not separated apart when expressed in a linear language, though it may not matter whether A is described before B or not. It is true that continuity remains continuous when projected on to the lower dimensional world. Hereupon, the psychological basis that ICS or PS can be a main character in human language is still in existence.

For the above reason, the author believes it is not wrong on the whole to think that human language has ICS or PS. It is said, however, that all existing languages in the world do not necessarily have phrase structure. It may be true as for their present forms. But from the development point of view, in the early times when only vocal sound were used as a language, it might have had phrase structure, and in the next stage when letters were used to represent sound elements, its structure was deformed little by little into the present form. The author supposes this change was brought about by the mechanism of human memory. That is, there are two kinds of memory, short-time memory (primary memory) and long-time memory (secondary memory). Languages in the form of vocal sounds are stored in primary memory in the first stage. But the capacity of primary memory is rather small and can hold but only for a short time, so if the spoken language has the crossed branches in syntactic structure, the previously stored information will be thrusted out before the related phrase appears. Therefore, that seems to be difficult for man to understand, and it can not be true

that such a difficult structure survives as a human language. On the other hand, when it is written in letters, it remains for a long time as it is before our eyes, since a sheet of paper plays the same role as long time memory does in the human brain. Then, even if the sentence is deformed, we can compare both forms, and recognize the deformed sentence as deformed one in relation to the original one. Therefore, we can accept the written language smoothly even if it is not PS, unlike the case of the spoken language. These deformed sentences, this time, are expressed in voice sounds, this spoken language may not have phrase structure, but it is accepted because the listener can imagine its original form with the aid of secondary memory. By the repeated process of this experience, transformed structure comes to generally accepted because of the adaptive ability of human beings. There may be some counter view which asserts that man can speak without knowing the written language. It is true, but in that case, the spoken language may have nearly phrase structure.

If it is reasonable to suppose that the present forms of written languages were caused by the modifications in a long time, the "transformational grammar" introduced by N. Chomsky can be said to represent the characteristic feature of human language activity, though transformation rules themselves are rather ad hoc. It is ad hoc because the transformations which are applied to the phrase structure language may not be necessary ones, but are only to give a strength or rhythm to the sentence, or clues for descrimination. Latin or other ancient languages may not possess phrase structure at surface structure, but they must have PS in their deep structure.

In any way, phrase structure is a good model of natural languages, if only the structure of language has something to do with the human

recognition or perception mechanism.

As mentioned above, the author believes that human languages have very regular rules in its basis which come from their physiological functions, and that English, as well as Japanese, has phrase structure in its very large parts. It is not necessarily easy to translate them mechanically because both English and Japanese have their own phrase structure. As in Fig.3.1.5, if a tree in one language differs from that of the other, then simple rule-by-rule translation can not be applied. These cases occur when units of recognition differ from language to language, and one-to-one correspondence in their units does not exist, even though the algorithm of recognition is the same in both languages. Therefore, if such case occurs, it is impossible to trans-

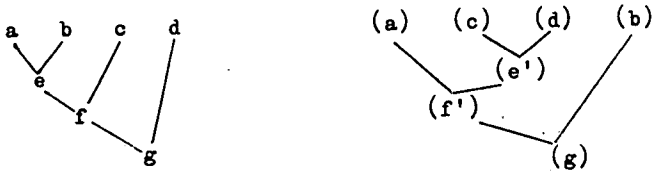


Fig. 3.1.5

late one sentence into another preserving one to one correspondence in words and syntax. But in the present state, to try far more flexible translation, concept to concept, is practically impossible because of the defects of our knowledge about our own languages. Then this paper will mainly be devoted to phrase structure, and investigate to what extent rule-by-rule translations are useful for English and Japanese.

3.2 Syntactic structure in English and Japanese

The manifest difference between English and Japanese in syntactic structure is word order and existence of particles in Japanese. It may be involved in the fact that the Japanese language is difficult to segment into word units, since KANA and KATAKANA and KANJI are used in mixture. This character prevents Japanese from being put into the computer and processed in it. But the problems of segmentation and alphabet system are not treated in this paper.

First, as for word order, it is apparently very different between English and Japanese. For example, in Fig.3.2.1, the word order in Japanese is almost reverse to that in English.

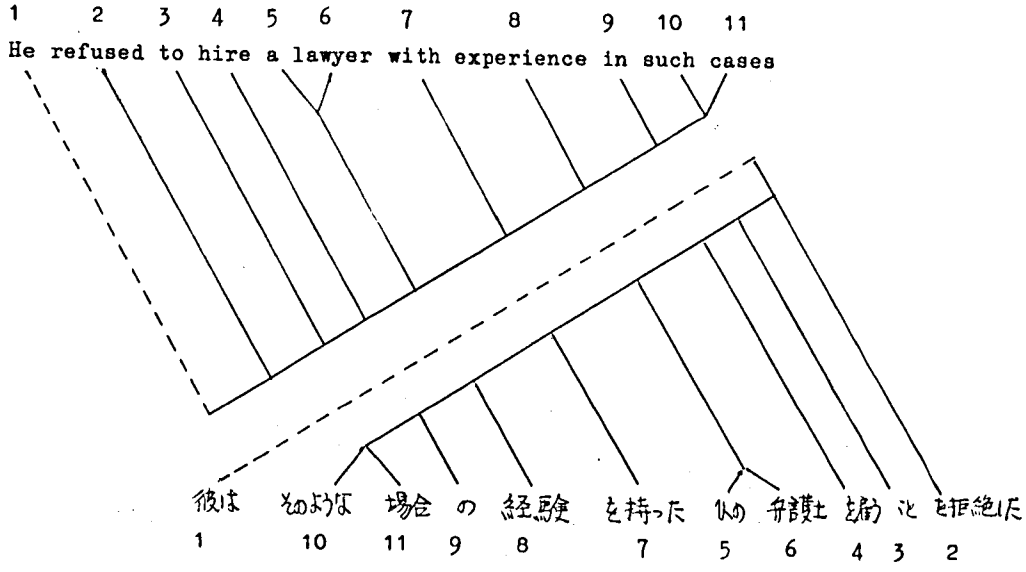


Fig. 3.2.1 Corresponding trees between English and Japanese structure.

But their tree structures show very beautiful relation hidden in both languages. That is, except for their subjective parts, both trees have very similar branches, and if one of them is put on another one,

turning inside out, both trees coincide quite closely. These trees shown in Fig.3.2.1 represents the typical structures of English and Japanese. In the English tree the right-side branches stretch up and the most extreme node is embeded in the extreme right side part, so that it is called right-recursive structure. On the other hand, in the Japanese tree the left-side branches extend upword, so that it is called left-recursive structure. In the typical example excluding the subject parts, the right-most part in English corresponds to the left most part in Japanese. Then to analyze English in order to translate it into Japanese, it is probably adequate or effective to begin parsing from the end of the sentence. In the reverse case, from Japanese to English, it may be good to begin from the top of the sentence.

Locally speaking, however, word order in English is not necessarily reverse to the Japanese one. They differ in noun phrase and verb phrase; in noun phrases, the words before the main noun, such as determiner or adjective or adjectival words, modify the main noun in such a manner that if they are situated nearer to the main noun, they are connected earlier than the farther words. As for the structure of modification in noun phrase, it is said that such structure as shown in Fig.3.2.2-a is not adequate, but all modifiers equally relate to the main word as shown in Fig.3.2.2-b. But the farther the modifier is apart from the main noun, the wider difference in intimacy reveals the con-

Fig. 3.2.2 Two structures of noun phrase; a-context free form b-true structure

cept of the modifier. Therefore, such a modification structure as in Fig. 3.2.2-a may be thought to be an acceptable one. As for the right-side modifier such as prepositional phrase or adjectival clause, they modify the noun phrase before them in the order of their appearance (Fig.3.2.3).

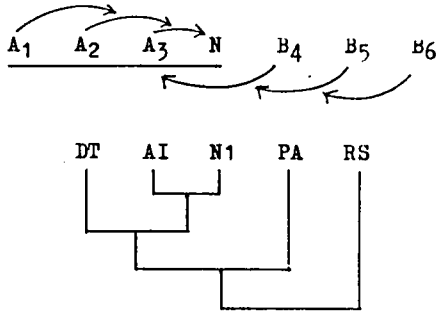


Fig. 3.2.3 Structure of noun cluster.

In verb phrase, the priority mode between the left-side and the right-side phrase is changed. That is, the right-side words or phrases modify the main verb before the left-side ones do. But the principal rule does not change, that is, the nearer they are to the main verb, the more intimate they are to the main verb (Fig.3.2.4). Usually left-side modifiers in verb phrase are adverbs, and the right-side modifiers are adverbs, noun phrase, prepositional phrase, and adverbial clause.

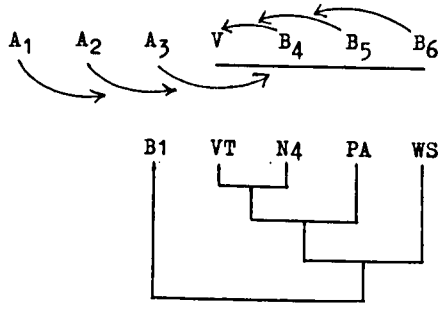


Fig. 3.2.4 Structure of verb cluster.

A few concrete examples are shown in Fig.3.2.5.

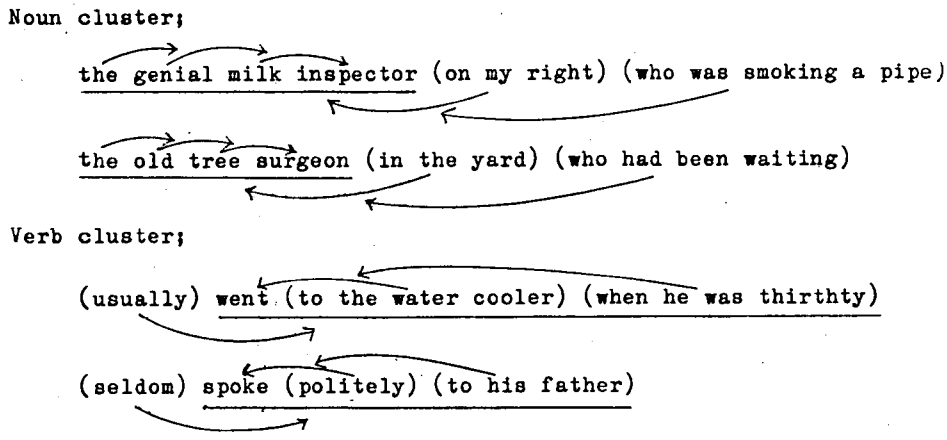


Fig. 3.2.5 Concrete examples of noun cluster and verb cluster.

The discussion mentioned above treats English structure alone. But the same modification principle can be applied to Japanese noun phrase and verb phrase. In Japanese, however, the direction of modification is one-way, that is, left to right (or up to down), and its principle is that the word nearest to the main word is connected most intimately to it quite in the same way as in English (Fig.3.2.6).

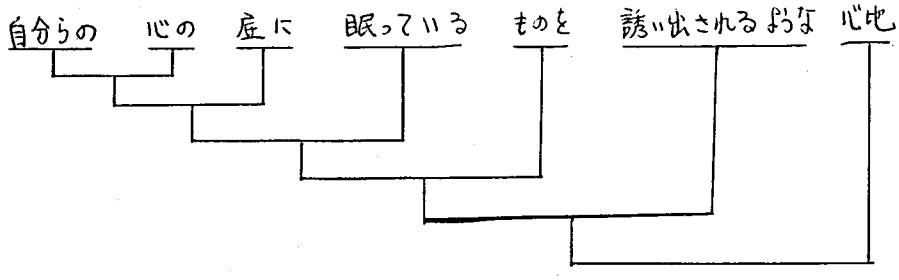


Fig. 3.2.6 Typical structure of Japanese.

It is very interesting that the order of modification in both languages is the same from the conceptual point of view, though the apparent

word order is different in both languages. That is, as shown in Fig. 3.2.7, if we put the number on each word according to the order in which the word modifies the main word, then the corresponding word or phrase in both languages have the same number.

a beautiful flower (in the vase)
2 1 3
(kabin no naka no) 1no utukusii hana
3 2 1

usually read (a book) (in the morning)
3 1 2
taitei (gozencyu ni) (1no hon) 0 yomu
3 2 1

Fig. 3.2.7 Modification order in English and Japanese.

This fact is not trivial, because modification order is independently determined in each language by using only syntactic information. This, however, does not mean that English speaking people read sentences forward and backward because the word order is different from the modification order in English.

The difference between English and Japanese in word order suggests that there are some differences in transforming tree or from two-dimensional relation to one-dimensional expression. However, the fact that the corresponding words have the same modification numbers tells us that the original materials in the real world which the phrase or clause describes are the same in both languages. This is analogously illustrated in Fig.3.2.8. That is, the images of A and B projected onto the x-axis and the y-axis are different in order, but the original relation

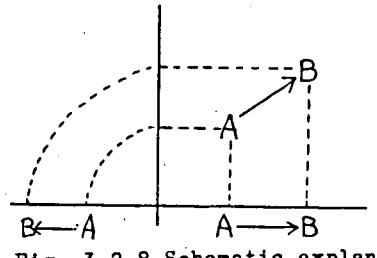


Fig. 3.2.8 Schematic explanation of word order in languages.

in two-dimensional plane is unique.

Though word order is different in both languages, it may be said to be a common feature in syntax that a verb and its object phrase are situated adjacent to each other in English and Japanese. This fact has no relation to mechanical translation. But if we look on the verb as an operator in mathematical formulae, then English and Japanese structure can be expressed like A + B and A B +. The former type is the normal expression form, and the latter looks like a polish notation which is convenient to be treated by the computer. Then Japanese may become a computer-oriented language, if it is well symbolized.

Now, the second point of difference between English and Japanese is that in English the role of the word is determined by its position in the sentence, but in Japanese the functional role is characterized by the particle, JYOSI. Therefore, in Japanese, word order is not so strict, and if only JYOSI's are predicted from the English structure, the word with adequate particles may be arranged almost arbitrarily so that mechanical translation becomes easier. If both languages have rigid structure, then rule-by-rule translation may be of no use because such occassion mentioned later in section 3.1 will occur. Japanese, however, is very flexible in a certain sense, so it is fairly possible to conform to English. But it is the question whether or not JYOSI's are correctly predicted by the English syntax, and how long a sequence of words is needed for this. Generally speaking, the number of branches coming into the node is usually two in immediate analysis, and about 99 percent of the sentences in English are said to be made up of two-branch nodes. Also in Japanese it is possible to parse two nodes into one, though the treatment

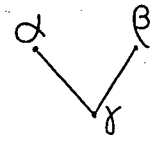


Fig. 3.2.9

Two-branch node.

of JYOSI is somewhat different from ordinary cases from the grammatical point of view. This two-branch analysis, however, is only effective to analyze one language as it is, and it is not enough when target language structure must be predicted by the analysis of the source language, as in the case of mechanical translation. For example, a phrase "the book which I read" is translated into "(WATASHI WA YONDA) HON" by two-

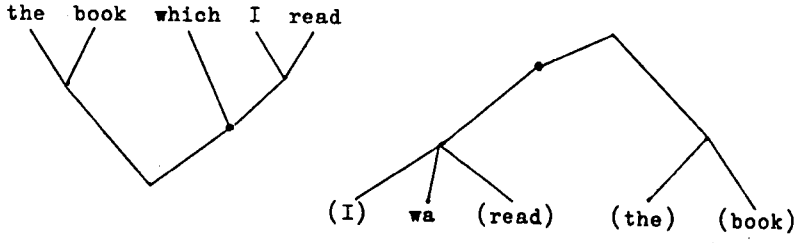


Fig. 3.2.10 Two-branch analysis.

branch analysis, because the embeded clause "I read" predict JYOSI "WA" in Japanese as in "WATASHI WA YONDA". But correct Japanese must be "(WATASHI GA YONDA) HON". The subject in a relative clause usually takes GA, and WA is used for a subject in the main clause. Then, for the above example, the tree looks as in Fig.3.2.11.

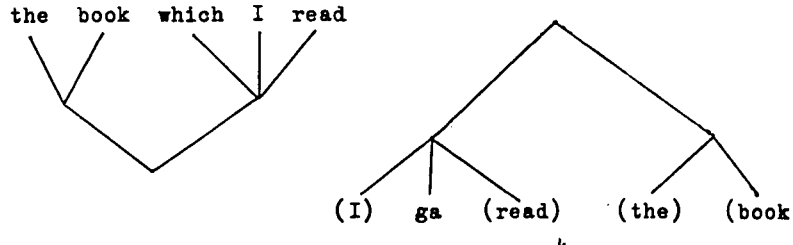


Fig. 3.2.11 Correct analysis by tree-branch rule.

This analysis tree may not be able to be called a grammatical one from the English syntactic point of view. This is a Japanese-language-oriented English analysis. This tree can be explained by mapping, as

shown in Fig.3.2.12.

Fig. 3.2.12 A transformation in Japanese tree.

The idea of mapping is introduced to rescue the weak points of context-free grammar; then it is not necessary if context-sensitive grammar is introduced. Context-sensitive rules, however, are somewhat troublesome to apply to the machine, so it is desirable to reduce them to context-free forms. This can be achieved by considering a somewhat long sequence of words as a left side of rewriting rules, that is, three (or more)- branch analysis. For example, context-sensitive rules (a), (b) are equivalent to (a)' and (b)'.

$$\alpha \cdot \beta \cdot \gamma \longrightarrow \alpha \cdot \delta \quad (a) \qquad \alpha \cdot \beta \cdot \gamma \longrightarrow \varepsilon \quad (a)'$$

$$\alpha \cdot \delta \longrightarrow \varepsilon \quad (b) \qquad \alpha \cdot \delta \longrightarrow \varepsilon \quad (b)'$$

Fig. 3.2.13 (a) and (b) is equivalent to (a)' and (b)'.

In general, phrase-structure grammar does not restrict the number of symbols in the right hand of a rewriting rule, but to construct a tree having two-branches at every node, two-symbol rules (n=2) must be intro-

$$\alpha \longrightarrow \beta_1 \cdot \beta_2 \cdot \cdots \cdot \beta_n$$

duced in context-sensitive form. If a node is allowed to have more than three branches, context-free rules can substitute for context-sensitive rules. At any rate it is necessary to take into considaration fairly wide-range context to insert appropriate JYOSI's in translated Japanese.

In summary, English and Japanese have, in its most parts, phrase structure, and both structures correspond to each other very well in tree diagrams, which can be expressed in the embeded box schema as in

English	Japanese
1+X(1+X(1+X(1+X(1+X))))	((((X+1)X+1)X+1)X+1)X+1

Fig. 3.2.14 Schematic structure of English and Japanese.

Fig.3.2.14. It is easily understood that English structure is right-recursive, and Japanese is left-recursive from the analogy of calculation of polyominals. To calculate an English-type polynominal, it is necessary to begin at the right part, that is, the most deeply embeded position. And if it is arranged in its calculation order, the result is a Japanese-type polynominal. Therefore, in English-Japanese translation it is advisable to begin analysis of English at the right most part toward the left part, unlike the ordinary English structure analysis, left to right.

3.3 Hierarchy in rewriting rules*

In the previous section 3.2, it is shown that there is the order of modification in noun phrase and verb phrase, here more detailed investigation will be given to each rule in syntax.

---

* Most of the symbols for parts of speech used in this section are listed in Table 3.4.1.1 (section 3.4.1). Other symbols will be easily understood by the analogy of the listed-ones.

First, in a tree of Fig.3.3.1, the rewriting rule AB ⟶ F is applied before FC ⟶ H . In this case we call AB ⟶ F is higher in hierarchy than FC ⟶ H , , or node F is of higher rank than H. That is, node x is said to be in higher rank than node y below it, if there are branches which connect x to y without passing root-point S. Otherwise x and y are not compared in rank, and said to be the same rank. In Fig.3.3.1, nodes F and G, or H and G are the same rank.

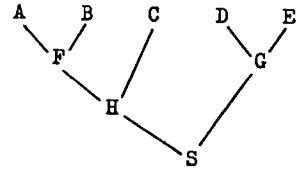


Fig. 3.3.1 A tree

Here the simple sentences which have typical structures are taken up to investigate hierarchies. In Fig.3.3.2, such a sentence is shown, and the hierarchies of rewriting rules are indicated.

(1) DT N1 ⟶ N4
(2) P1 N4 ⟶ PA
(3) N4 PA ⟶ N4'
(4) VT N4' ⟶ VE
(5) N4 VE ⟶ SS

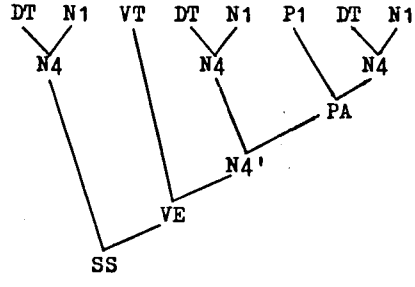


Fig. 3.3.2 Elementary hierarchies in rewriting rules.

If the hierarchies are ignored, the example shown in Fig.3.3.3 is analyzed in a wrong way, because there are N1 PA ⟶ N4 and N4' PA ⟶ N4 , and analysis is begun at the end of the sentence, N1 PA ⟶ N4 is applied before DT N1 ⟶ N4', then it is as shown in Fig.3.3.3. If the hierarchies indicated in Fig.3.3.2 are considered, DT N1 ⟶ N4' is applied before N1 PA ⟶ N4 , and the correct tree can be gotten, as

shown in Fig.3.3.2.

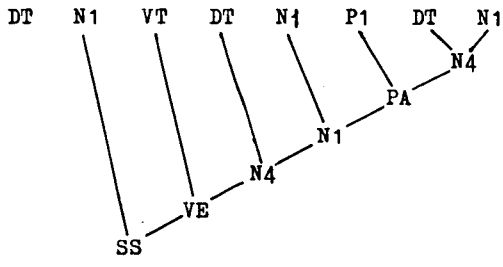


Fig. 3.3.3 A wrong analysis by ignoring hierarchies.

These hierarchyies, however, depend on the direction of analysis, that is, left-to-right or right-to-left parsing. For example, by the right to left analysis the sentence "She has books in her hand" is correctly analyzed without any hierarchy, but in the left to right analysis, the passing is blocked, as shown in Fig.3.3.4.

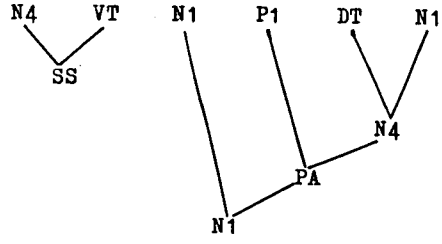
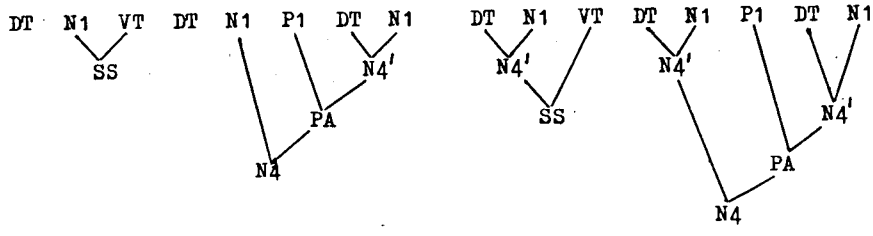


Fig. 3.3.4 An imcomplete tree by left-to-right parsing.

The reason why rewriting rules must be classified into several hierarchies is that in ordinary sentences nouns with or without determiners or adjectives are modified by prepositional phrases, or verbs appear without their objects, then there must be prepared such rules as

N1 PA → N4 N4 PA → N4' DT N1 → N4' AI N1 → N1 N1 VT → SS
VT N1 → VE N1 VE → SS

etc., and that these rules, if applied in the same hierarchy, lead to the wrong result in case of N.V D N1 VT DT N1 P1 DT N1 regardless of the parsing direction, as shown in Fig.3.3.5.



right-to-left parsing left-to-right parsing

Fig. 3.3.5 Incomplete analysis by ignoring hierarchies.

Then at least five hierarchies mentioned above are needed. If these hierarchies are taken into consideration, the direction of parsing for simple sentences is free, but right-to-left analysis is preferable in reducing the number of procedures for parsing.

In the previous example (Fig.3.3.5), such rules as N1 PA → N4 , N4' PA → N4 , VT N1 → VE , N1 VT → SS , N1 VE → SS etc., are put into the same class and given the same hierarchy, so that they interact on each other. But there are several rules which do not interact on each other. If the parsing direction, right to left, is taken into consideration, then it may be possible to reduce the number of hierarchies. Then, as the first step, all rules are divided into two classes so that each class does not include any intersecting rules.

(I) DT N1 → N4' P1 N4' → PA N4' PA → N4 P1 N1 → PA
VT N4 → VE N4' VE → SS N4 VE → SS

(II) N1 PA → N4 VT N1 → VE VT N4' → VE N1 VE → SS
N4 VT → SS N1 VT → SS

They are (I) and (II). In each class all rules have the same hierarchy, but the rules in class (I) are in a higher hierarchy than those of class (II). In this classification all simple sentences can be analyzed correctly by right-to-left parsing. " DT N1 VT N1 P1 DT N1 " (D N V N P D N), for example, takes the form as in Fig.3.3.6, where the wave-line branch shows that class (II) rules are applied. In class (I),

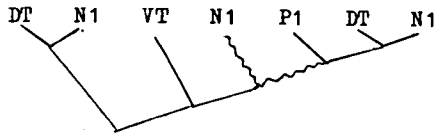
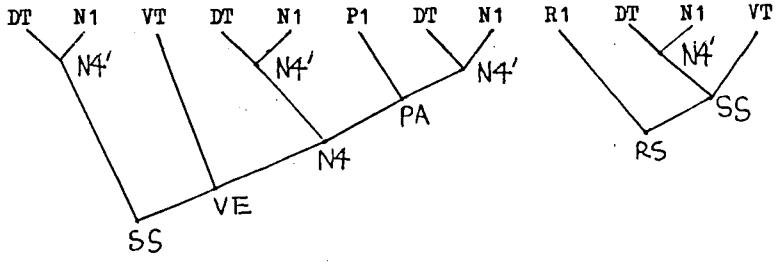


Fig. 3.3.6 Analysis using (I) and (II).

there are typical case rules, and in class (II) somewhat irregular type rules. To put class (I) rules in a higher hierarchy than class (II) makes context-free style rule (II) equivalent to a context-sensitive one, because the context for rule (II) is investigated by the rule (I) which is applied to it before them.

When only simple sentences are concerned, it is enough to divide them into class (I) and (II), as mentioned above. But to translate also complex sentences, as well as simple sentences, some further consideration is needed. If simply we add R1 SS → RS , N4' RS → N4 , and VT N4 → VE to class (I), and N1 RS → N4 to class (II), then "D N V D N P D N R D N V", for example, becomes as below.



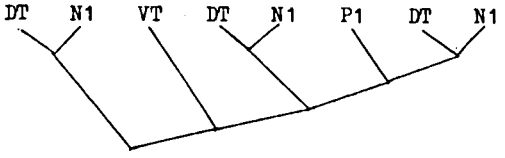
This result is not correct. That is because the main part of this sentence is a standard form, and is analyzed only by using class (I) rules, but the right part of the relative pronoun, (DNVN), belongs to the irregular sentence, so it is left unanalysed. This fact, however, depends on the algorithm of applying rule (I) and (II). To tell the truth, the rule must be applied at first to the simple sentence embeded in the right most part, and then the same processing must be applied recursively to the result of the previous procedure. But generally, it is very difficult to find out a simple sentence in a somewhat complex structure. For example, in " /I will show you /some results/ obtained by the digital computer ", there are several possibilities to cut. Further, if there are inserted phrases or clauses, as in "I will show you, here, some results obtained by the digital computer", then verb phrase beginning with "obtained" may be looked on as predicate of "some results".

Then classification of rules is slightly changed, as shown in (I)' and (II)'. Class (I)' includes mainly noun-type rules and preposition-

(I)'	DT N1 $\longrightarrow$ N4'	P1 N4' $\longrightarrow$ PA	P1 N4 $\longrightarrow$ PA
	P1 N1 $\longrightarrow$ PA	N4' PA $\longrightarrow$ N4	N4' RS $\longrightarrow$ N4
(II)'	VT N4' $\longrightarrow$ VE	VT N4 $\longrightarrow$ VE	VT N1 $\longrightarrow$ VE
	N4 VE $\longrightarrow$ SS	N4' VE $\longrightarrow$ SS	N1 VE $\longrightarrow$ SS
	R1 N1 VE $\rightarrow$ RS	R1 N4' VE $\rightarrow$ RS	etc.

al phrase, and class (II)' includes verb-type rules and sentence-form rules. By applying the class (I)' rule first, a noun modified from the left side (determiner, adjective) becomes noun phrase before it is connected to a verb, so such a case as shown in Fig.3.3.5 does not appear. The previous example which is analyzed incorrectly by class

(I) and (II) become as follows;



As illustrated above, it is only necessary to classify all rules into two groups, (I)' and (II)', though in the above class (I)' and (II)' only necessary rules are illustrated. This is because right-to-left parsing is pre-supposed. The meaning of "right-to-left parsing", however, must be described. In the above parsing, in the first step, a limited length string (substring) of parts of speech in the given sentence is taken up in turn beginning at the end part of the whole string, and each time it is compared to the rule in class (I)' and processed if necessary, and in the second step, class (II)' rules are used in the same way. Therefore each rule in class (I)' or (II)' is in the same hierarchy. If one rule is picked up from class (I)' or (II)' and compared to every substring and processed according to the rule, it is equivalent to the case of the totally ordered rules. The totally ordered rules are effective but it takes increasingly more time to analyze as the number of rules increase. In this paper it is one of the aims to reduce processing time, so the total order system is not adopted.

The above discussions are based on the hypothesis that English structure is right-recursive, but in actual situations, there often appear not simply right-recursive sentences. For example, in the sentence

> " In many cases [difficulties which appeared to arize from semantic or syntactic problem] were rather easily cured by slight modification in structure."

the subject part bracketed by [ ], which contains a clause in it, appear before the predicate. So if the parsing is carried out from right to left in one way, the predicate of the main sentence is apt to be parsed with the noun phrase embeded in the adjectival clause in the subject parts before the subject part is correctly parsed (Fig.3.3.8). As for the solution of these cases, it is one method to modify slightly the original sentence itself, just as explained in the above sample sentence, or another is to divide class (II)' rules into two more parts, $(II)'_1$ and $(II)'_2$, giving $(II)'_1$ a higher rank than $(II)'_2$.

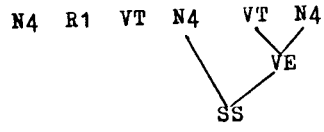


Fig. 3.3.8
A wrong analysis

$(II)'_1$ $VT \cdot N4' \rightarrow VE$, $R1 \cdot N1 \cdot VE \rightarrow RS$

$(II)'_2$ $N1 \cdot VE \rightarrow SS$, $N4' \cdot VE \rightarrow SS$, $N4 \cdot VE \rightarrow SS$

As for modification, it is simple to insert a comma between subject part and predicate part as a barier , in the above case between "problem" and "were".

To sum up the discussion in this section, first, if in rule-by-rule translation scanning one direction is always adopted, then rewriting rules in context-free form must be classified into several hierarchies. Second, the hierarchy of rules depends on the scanning direction, right to left or left to right. Third, through possession of a proper hierarchy, context-free rules function as if they were context-sensitive.

The hierarchy of rules depends on the symbols or parts of speech

which constitute the rules. For example, in DT·N1 $\rightarrow$ N4', two symbols N1 and N4' are used to represent necked noun and modified noun. But when N1·PA $\rightarrow$ N4 and DT·N1 $\rightarrow$ N1 are separated, symbol N4' is unnecessary. Therefore, to introduce new symbols is equivalent to classifying rules into smaller classes, that is, to increase the number of hierarchies. Increase in kinds of symbols, however, leads to the increase in the number of rules. On the contrary, decrease in kinds of symbols makes it difficult to analyze English and to predict Japanese structure. So it is a very important problem to choose appropriate symbols so that they represent our grammatical concept as well as possible and characterize both English and Japanese.

3.4 Classification of parts of speech

In the case of human translation one not necessarily needs grammatical knowledge about parts of speech, but has only to know the meanings of words and functions of a comparatively small number of words. This is because he knows the real objects which the words signify, and relations between words in a sentence can be recognized through the actual relation among objects themselves; therefore, such information as noun, verb, or adjective, etc. are not necessary for him. On the other hand, when we are to decipher a cryptogram or to read unknown languages, we can not get concrete images from the words in it, then some information substituting for meanings is needed. The part of speech plays its role. In English-Japanese mechanical translation, the part of speech gives a clue of finding the departure point

in syntax between the source and the target languages, by which it is possible to exchange word orders and insert adequate JYOSI(Particle) in the target language. For this purpose, traditional classification of parts of speech, as in descriptive linguistics which is human-oriented grammar and treat only one language independently, is not useful. Also such a classification method as C.C. Fries', where frame sentences determine the word class as a set of words which can occupy the same position in the frame sentence, may lead superficially to the same result as will be explained in this paper, but it is not applicable to mechanical translation in its fundamental philosophy, especially in the case of treating such languages from different language-families. Even the parts of speech of the source language must be classified by taking into consideration the structure of the target language. The term "part of speech", however, is used in this paper as having an equivalent meanings with an identification symbol, and not used as a traditional concept.

The fundamental attitude for word classification is illustrated by a few examples.

In the next two sentences (1) and (2), phrases led by "to" work differently in syntactic structure, and so the Japanese word for "to" also differs in Japanese.

(1) I go to school. watashi wa gakko e iku.

(2) I eat to live. watashi wa ikiru tameni taberu.

It is only determined by the context in what way the word "to" works. Therefore, it must be distinguished from ordinary prepositional words, though it may work as a preposition.

In the next phrases, the adjectival words "interesting" and "useful" play the same role in the sense that both words occupy the same

(3) This is an interesting book. (3)' This book is interesting.

(4) This is an useful book. (4)' This book is useful.

position in the sentence. Therefore "interesting" and "useful" may be put into the same class according to the ordinary classification. The corresponding Japanese for above examples, however, are different, as below,

$(3)_J$ kore wa omoshiroi hon de aru. $(3)'_J$ kono hon wa omoshiroi.

$(4)_J$ kore wa yuekina hon de aru. $(4)'_J$ kono hon wa yueki_ de aru.

The points of difference are concerned with the inserted particles. That is, the word "interesting" requires the same particle "-I" in both cases (3) and (3)'; on the other hand, the word "useful" requires different particles " -DE" and " -NA" because of its contextual differences. If both words are given the same symbol, they can not be translated correctly because there is no clue to distinguish between them from the syntax. Therefore, they must be separated form the Japanese translation's point of view.

In the other examples (5) and (6) below, if "in" and "of" have the translation "NO NAKA" and "NO", example (6) needs no particles for

(5) an apple in the basket kago no naka no ringo

(6) a friend of mine watasi no __ yuujin

preparational phrase in Japanese. If "of" is given the same symbol, then the translation (6) becomes "**watasi no no yuujin**". Therefore, prepositions "in" and "of" must be put into different classes from each other.

As shown above by a few examples, word classification must be performed by considering the correspondence in word order between both languages, the necessity of particle insertion, the inflection of appropriate words in translation, and also the syntactic character of English itself.

First, English words are divided into two categories, that is, form class and function class. Form class includes most of so-called nouns, verbs, adjectives, and adverbs. This class is almost the same as the ordinary form class. A word which is named as verb, however, does not necessarily construct the predicate; it may be of adjectival use or even of noun use in the given context. The word "context" means a sequence of some symbols. The part of speech (or word class or symbol) is used to give special features to contexts, connects words and phrases and makes it easier to translate English into Japanese. The other words which are not in the form class belong to the function class. A word which has the possibility of working both as a form-class word and as a function-class word, is given a special symbol and put into the function class. In the form class, each word can belong to at most two sub-classes. The detailed explanations are given below.

3.4.1 Form class

The form class is divided into four sub-classes, that is, noun

class, adjective class, verb class, and adverb class. The criteria for sub-classification follows the common usage.

Noun : N1

The noun class has no sub-class, because there is no special structual difference to distinguish among abstract nouns, concrete nouns, and material nouns. Pronouns, however, belong to the function class. As for proper nouns, it is convenient, if possible, to distinguish them from the other nouns, but in the present system input texts are typed, using only capital letters without putting any special marks to proper nouns, so it can not be distinguished automatically. Singular nouns and plural nouns can be in large part recognized by their inflection endings automatically, but such distinction is almost of no use because the number of a noun is not explicitly represented in Japanese syntax, and also because the case of concordance of noun and verb in number is in most cases redundant. And when it is not redundant as in the case where the antecedent of a relative pronoun is situated apart, it is too complex to treat by rule-by-rule translation. Nevertheless, sometimes the plural form is useful in finding the end point of a noun phrase; for example,

```
"  In   many   cases   difficulties   which   ....  "
   P1   AO    N1          N1          R1     ....
```

in the above example, if plural information is considered, then its structure can be looked on as " in (many cases) difficulties which ", because the plural-form noun usually does not modify the word behind it. On the contrary, if plural information is ignored, then the structure of the above case becomes " in (many cases difficulties) which ", because the "noun + noun = noun" pattern often appears.

To distinguish numerals or signs from ordinary nouns (N1) is useful in case of date expressions or equations in mathematics. Because such sentences as shown below which have different structures in meaning are looked on as if they were the same in syntactic structure, if there is no distinction between nouns and numerals.

It	happens	in	August	28,	1967.
N3	VI	P1	N1	N1,	N1

I	will	give	you	some	money,	Jack.
N4	VA	VD	N4	AO	N1	, N1

Numerals, however, are used in many ways other than in date expression, and so it is difficult to judge whether they represent dates or not. This problem is raised by the co-existence of scientific texts and rather conversational texts. If the latter texts are ignored, then it is possible to treat date expressions or equations without further classification of noun classes.

Adjective : (AI, AN, AO)

Adjectives are classified into three sub-classes according to their endings or conjugation suffixes when they are translated into Japanese. That is, class AI contains such words as require particle "-I" when they modify nouns, class AN requires "-NA", and class AO requires "-NO". For example,

a beautiful flower AI	utukusi-<u>i</u> hana
a pretty flower AN	kirei-<u>na</u> hana
a brown flower AO	kassyoku-<u>no</u> hana

These words which belong to class AN and AO are also used as noun-

like words, and in that case their conjugations differ from those of adjectival use. Therefore, it is not a good way to store in the dictionary their equivalents in Japanese with inflection particles. The stem forms which are registered in the word dictionary and their conjugated form are shown below.

English word	class	Japanese word	Example English	Example Japanese
small	AI	tiisa	It is small.	sore wa tiisai.
easy	AI	yasasi	It is not small.	sore wa tiisakunai.
disobedient	AI	jyujyundena	a small apple	1no tiisai ringo
abnormal	AN	ijyo	He is abnormal.	kare wa ijyo de aru.
cynical	AN	reisyoteki	He is not abnormal.	kare wa ijyo de aranu.
strange	AN	kimyo	an abnormal man	1no ijyona otoko
final	AO	saigo	it is final.	sore wa saigo de aru.
handwritten	AO	tegaki	It is not final.	sore wa saigo de aranu.
lossless	AO	musonshitsu	a final notice	1no saigono tsukoku

Such particles as "I", "NA", "NO", and "DE" are instructed by rewriting rules.

The comparative or superlative degree (RA) of class AI, if it is distinguished by different symbols from the positive degree, makes translation into Japanese more delicate. But from the syntactical stand point, they are not necessarily effective because they do not differ from ordinary adjectives in word order both in English and Japanese.

Verbs VI, VT, VD, GI, GT, GD, PI, PT, PD

Except some verbs which belong to the function class, verbs are classified into three sub-classes which correspond, on the whole, to intransitive verb, transitive verb, and dative verb. In the ordinary grammar it is said that transitive verbs take objects, and intransitive verbs take complements, and dative verbs take double objects; however, the distinction between objects and complements depend on the meanings of the words. Here, verbs are classified according to the particles in Japanese which the verbs require, when an English sentence in which nouns or their equivalents are situated just behind the verbs is translated into Japanese. That is, as shown below,

He accepts the opinion. (VT)	kare wa sono iken o ukeireru.
He becomes a chief. (VI)	kare wa 1no syunin ni naru.
He gives her a flowere. (VD)	kare wa kanojyo ni 1no hana o ataeru.

When a sentence which has "...V·N^1 N^2 " is translated into Japanese, if the verb requires particle "O", then the verb is named VT, and if it requires "NI", then it is named vI, and if it is translated as "Nl = N2ヲV" or "NlヲN2 = V" , then it is named VD. There are, however, several English words which can not belong to any class by this criteria; for example, those words which always appear solitarily without any noun-like word, or such a word as "seem" which requires somewhat different particles. These words are put into class VI. The syntactic priority-order is given to these three classes. That is, the first order is given to VD, the next to VT, and the last to VI. Therefore, if a word has a possibility of belonging to more than one

class, the word is classified into the highest priority class.

Each class VI, VT, or VD is divided into three sub-classes according to the conjugational forms of English words. They are the present form (VI, VT, VD), past (participle) form (PI, PT, PD), and ing-form (GI, GT, GD). The past participle form can not be distinguished from the past form only by using morphological information except in case of some irregular verbs, so they can not but have the same symbol. But their roles in the sentence can be distinguished by the context in most cases. For example, in the sequence of symbols "N4 PH PT DT N1" the word named PT is used as a past participle because a word (VH) just before it belongs to the "have-class" and PH PT construct a past perfect expression. About the more complex cases such as "N4 VT DT N1 PT P2 N1"(he wants a result obtained by computer) , they will be explained in the next section.

It may seem to be unnecessary to distinguish transitive verbs from dative verbs because their function can be determined by the context. In fact, if a verb appears in the context V N4 N4 , this word is probably a dative verb, and if it appears in V N4 , then it may be a transitive verb. But sometimes there may appear such cases as shown below where the first noun $N4^1$ is the true object but the second noun $N4^2$ is a subject for the following parts; therefore, it is not reliable to judge such a verb as dative from the mere sequence of parts

$$(\ldots\ldots\ VT\ \ N4^1\ )\ \ (\ N4^2\ \ VT\ \ldots\ldots\ )$$

of speech. If dative verbs are separated from transitive verbs, the syntactic information is increased and so the parsing becomes more probable even in such ambiguous cases.

TABLE 3.4.1.1 Classification of English words.

Symbol	Mnemonic name	Example
Form class		
N1	Noun	book, books, air, 23, Jack, Monday, etc.
AI	Adjective	beautiful, ambiguous, big, cute, etc.
AN	"	profitable, stupid, traditional, etc.
AO	"	native, normal, passive, sensory, etc.
RA	" comparative	best, better, larger, biggest, etc.
B1	Adverb	mainly, however, now, then, thus, etc.
VD	Dative verb	give, makes, etc.
PD	" past-form	gave, given, made, etc.
GD	" ing-form	giving, making, etc.
VI	Intransitive verb	go, coincide, obey, etc.
PI	" past-form	went, gone, lived, etc.
GI	" ing-form	going, occuring, running, attending, etc.
VT	Transitive verb	avoid, compute, hurt, kills, love, etc.
PT	" past-form	invited, taught, thought, supposed, etc.
GT	" ing-form	speaking, reading, reducing, etc.
Function class		
DT	Determiner	a, an, the, my, their, each, such, etc.
DY	"	every, etc.
N3	Pronoun	it, this, these, those
N4	Noun	I, he, me, us, she, etc.
P1	Preposition	at, in, on, under, beside, between, etc.
P2	"	of, with, across, by, per, etc.
P4	"	for, since, after, etc.
AS	"	as
TO	"	to
C1	Conjunction	and, or, but, nor
C2	"	if, though, because, so that(idiom), etc.
--	"	- (hyphen)

TABLE 3.4.1.1 Classification of English words. (continued)

Symbol	Mnemonic name	Example
Function class (continued)		
VA	Auxilially verb	will, shall, can, may, must, etc.
PU	" past-form	would, should, could, might, etc.
VF	Do verb	do, does,
PF	" past-form	did, done
GF	" ing-form	doing
VH	Have-verb	has, have
PH	" past-form	had
GH	" ing-form	having
VL	Be verb	be, is, am, are
PL	" past-form	was, were, been
GL	" ing-form	being
HH	Adverbial	there, here
QQ	"	yes, no
XX	"	let, please
EE	"	not
R1	Relative pronoun	who, which
R2	"	whose
R3	"	whom
TH	"	that
W1	"	where, why
W2	"	how
WN	"	when
WH	"	what
,,	Punctuation mark	, (comma)
((	"	((left-bracket)
))	"	) (right-bracket)
''	"	' (apostrophe)
¥¥	"	? (question mark)
!!	"	! (exclamation mark)
#Δ	"	(beginning mark of sentence)
Δ#	"	• (end mark of sentence)
**	"	(sentence boundary)

Adverbs : (B1)

There is no sub-classification of adverbs in the present system. But there are some words which are difficult to classify as adverbs, because they are used both as adverbial and noun like-words. For example, in the sentence "I do not go to school every Sunday.", the phrase "every Sunday" is used as adverbial. Then, is the word "Sunday" an adverb? The answer must be "yes" in this case from the semantical point of view. But in the next sentence "Sunday is a public holiday", the same word "Sunday" is clearly used as a noun. Where does this difference come from? It is difficult to answer such a question only from the syntactic stand point. One of the explanations may be as follows. There may exist adverbial nouns which have meanings concerning time or place, and if they are modified by such words as "every", "later", etc., they work as adverbs. Otherwise, they function as nouns. Therefore, if such words as "every" and "later" are given a special symbol which represents only such words, and adverb-like nouns are distinguished from ordinary nouns, then it may be possible to correctly translate the above two sentences. In the scientific readings, however, the adverbial use of such words is comparatively rare so that in this paper they are included in the same class as noun. Nevertheless, more detailed classification of adverbs may be effective in machine translation.

In English, adverbs or adverbial words can be put in almost any position between phrases; this feature makes it difficult for mechanical translation to process all syntax by the rule-by-rule method.

The form class mentioned above has, in total, fifteen sub-classes.

3.4.2 Function class

Determiner (DT)

The words which belong to the determiner class modify nouns on the right side of them, but unlike adjectives they are not modified by other words from the left side. In other words, determiners work as initial words of noun phrases. Some examples of determiners are shown below. DY is a sub-class of determiners, and words belonging to this class construct adverbial phrases when they modify some nouns, for

DT : a, an, the, such, all, any,
each, his, my, its, their, etc.

DY : every

example, "every day" or "every time". But in the present system the noun class is not classified into sub-classes by considering semantic aspects; therefore, it is erroneous to regard a concatenation of DY and noun simply as an adverbial phrase. There are several exceptional expressions with determiners; for example, "such a" or "many a" etc. In these cases, "a" is modified by an other word, so strictly speaking, it can not belong to the determiner class by definition. Nevertheless, these phrases "such s" and "many a" can be looked on as if they were one word, so they are to be substituted by a single word belonging to the determiner class at the idiom-processing stage in the mechanical translation system.

Several words, such as "all" and "each", work rather irregularly, as compared with other words; for example,

They all went to the station.

They cost six pence each.

Therefore, such a word must be distinguished from ordinary determiners, though sometimes it is possible to determine its role from the context only.

Pronouns : (N3, N4)

Pronouns are divided into two classes N3 and N4.

N3 : it, this, these, those

N4 : I, he, we, him, they, their, etc.

These words are not modified from the left side, and except some of N3, they do not modify other nouns. The main difference between N3 and N4 is that in the next sentence it is differently translated into Japanese,

It is very curious that he did not know it.
RS
kare ga sore o siranakkata koto wa taihen kimyo de aru.
RS

He believes that languages have phrase-structure.
N4 RS
kare wa gengo ga ku-kozo o motu koto o sinjiru.
N4 RS

depending on whether the first position is occupied by N3 or N4. The translation word for an English word named N3 does not appear in Japanese, it is substituted by a phrase named RS.

Functional Verbs : (VA, VF, VH, VL, PA, PF, PH, PL, GF, GH, GL)
U

Such words as shown below play characteristic roles in indicating the aspect of verbs and also making sentence structure (interrogative or negative sentence).

VH :	have, has	GH :	having	PH :	had
VF :	do,does	GF :	doing	PF :	did, done
VL :	is, am, be, are	GL :	being	PL :	been, was, were
VA :	will, can, may,			PU :	would, could, might,

Sometimes VH class words (have) or VF class words (do) are used as if they were transitive verbs. In such cases, however, their function can be recognized by the context, for example, in the sequences "...N4 VH N4..." and "...N4 VF N4..." , VH or VF are used as transitive verbs, and in the sequence "...N4 VH PT N4..." or "...N4 VF EE PT N4..." VH and VF are used as auxilialy verbs. The word "may" or "will" is sometimes confused with the word which means "month" or "intention". To avoid these confusions, they must be, as a rule, given special symbols different from those for other VA class words, but it is not impossible to decide by the syntactic context alone in which way the words are used in the given sentence. Therefore, they are included in VA class.

Preposition : (P1, P2, P3, P4, ...)

Preposition or preposition-like words are divided into several classes according to their roles and the kinds of particles which they require.

P1 : in, on, below, between, under, etc.

P2 : of, across, against, by, into, from, etc.

P1 and P2 classes include prepositions which do not work as conjunctions. Words in P1 group need particles when they modify nouns or verbs as prepositional phrase; on the other hand, words in P2 do not need particles, as shown below.

He lives in a large house. (P1) kare wa 1no ookii ie no naka ni sumu.

This book was written by him. (P2) kono hon wa kare ni yotte kakareta.

Several words serve as both preposition and conjunction, and they are given special symbols, because their rules are only determined in the course of analysis. For example, if a word in P4 class appears in

I can not agree with you , for the proposal is too severe.
P4 N4 VE
B2

the context "...P4 N4 VE..." , it may be probable to infer that it serves as subordinate conjunction.

Other words which work like prepositions or conjunctions are "to" and "as" or "like", etc. As for "to", it differs from other prepositions by the fact that it can take the root-form of a verb just behind it, and its translation, as well as structure, in Japanese is very different according to whether it takes noun phrase or verb phrase as an object. It may be, however, possible to identify by the context the case where "to" is used as part of a to-infinitive even if "to" is included in P1 class, because it is the only word that takes the verb as its object. But in complex sentences it is convenient and reliable to distinguish

it from other prepositions. The function word "as" behaves in a more complex way, and must be given a special symbol to recognize its role in the given sentence.

Relative Pronouns and Interrogative Pronouns.

Relative pronouns or interrogative pronouns play many roles and behaves in a somewhat complex manner, and so it is difficult to give them appropriate translations. Perhaps it is necessary to give a different symbol to each word to characterize their syntactic aspects. But in scientific literatures, interrogative sentences seldom appear, so several words can be grouped according to their typical use. The classes are shown below. "THAT" is a multic function word, and constructs one class.

R1 :	who, which, whoever, etc.	W1 :	where, why, etc.	WH :	what
R2 :	whose	W2 :	how	WN :	when, etc.
R3 :	whom	TH :	that		

Conjunction (C1, C2, --.)

There are three kinds of conjunctions, that is, co-ordinate conjunction(C1), subordinate conjunction (C2), and hyphen (--). A hyphen is not a conjunction in an ordinary sense, but from the syntactical view point, it behaves like a co-ordinate conjunction. The different use of hyphen and co-ordinate conjunction appear in such expressions as "time-sharing system" or "inflected-form".

C1 : and, but, or, etc.

C2 : if, because, though, etc.

The words which belong to C2 class govern the sentence structure and make it an adverbial clause. The syntactic character of C1, C2 and hyphen is as below,

boys and girls N1 C1 N1	:	syonen sosite syojyo N1 C1 N1
if you go C2 N4 VI	:	anata ga iku nara N4 VI C2
punched - card PT -- N1	:	sankoosareta - kaado PT -- N1

Miscelaneous

It is necessary to distinguish the word "not"(EE) from ordinary adverbs, because its translation affects the conjugation of a verb. The word "there" and "here"(HH) are also separated from other adverbs because they are used in a special way as "There is a book on the desk". The translation of the example becomes "TUKUENO UENI HON GA ARU", and the correspondence of English to Japanese in this case is a very special one; therefore, such a word must be given a special symbol to recognize the characteristic feature of English syntax. It is not necessarily effective to distinguish such words as "yes" or "no"(QQ), but sometimes in high school texts they are useful. To such words as "please" and "let"(XX), a special symbol must be given to treat imperative sentences. These words (EE, HH, QQ, XX) can be looked on as belonging to a sub-class of adverbs, but they are very different in use from the ordinary adverbs.

Punctuation mark

Punctuation marks are as important as the parts of speech. These marks are considered at the same level as ordinary parts of speech.

The initial mark (#Δ) is useful to specify a verb at the beginning of a sentence, and to recognize the syntactic feature of imperative sentence or participial construction. The final mark (Δ#) which corresponds to the period guarantees that there are no more words to the right side of it. The intermediate mark (**) is used to separate two sentences so that the rule may not be applied to the two successive sentences when many sentences are translated at a time. Commas (,,) plays a very important role in segmenting phrases or clauses, though sometimes they make sentence analysis rather difficult. The brackets (((,))) are also used at the same level with nouns or verbs. Apostrophies ('') and question mark (¥¥) are also considered as parts of speech.

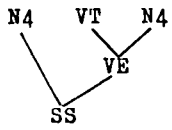
3.4.3 Cluster symbols

The above mentioned symbols are functional ones, and there are in total 41 symbols. These symbols which belong to the form class and the function class are, in principle, given to single words when each word is looked up in the word dictionary. On the other hand, several symbols are necessary to represent phrase, clause and some sequence of symbols which correspond to the node names in the syntactic tree. In this paper such symbols are called "cluster symbols", but they are

TABLE 3.4.3.1 Classification of clusters.

Symbol	Mnemonic name	Example
N1	Noun phrase	AI + N1, N1 + N1,
N4	"	DT + N1, N4 + RS, N4 + PA,
PA	Prepositional phrase	P1 + N4, P1 + GT + N1,
PB	"	P2 + N4, P2 + PI + N1,
VE	Verb phrase	VT + N4, VI + AI, VL + N4,
PE	" past-form	PT + N4, PH + N4,
GE	" ing-form	GT + N4, ,, + GE + ,,
B2	Adverb	,, + B1 + ,, , ,, + B1 + △#,
L1	"	TO + N4,
L2	"	TO + VE,
RS	Relative clause	R1 + N4 + VE, R2 + N1 + VE, R1 + VE,
WS	"	W1 + N4 + VE,
SS	Sentence	N4 + VE, N3 + VE + RS, SS + ,, + SS,
C3	Conjunctive	,, + C1
C4	"	,, + N1+C3
UU	Top word	#△ + HH
FP	"	#△ + PF
LP	"	#△ + PL
A.	"	#△ + VA
H.	"	#△ + VH
HP	"	#△ + PH
AP	"	#△ + PU
F.	"	#△ + VF
L.	"	#△ + VL
WI	"	W2 + AI
WA	"	W2 + AN
**	"	** + SS, ** + **

often used on the same level as ordinary parts of speech. Cluster symbols must reflect the features of constituents, that is, it is desirable to be able to guess the structures of sub-trees growing on the node by the symbol given to the node. It is apparent that if the number of cluster symbols are allowed to increase, it is easy to guess the structure of a sub-tree by the node name. Increase in the number of symbols, however, causes the increase in the number of rewriting rules which are made of concatenations of several symbols, and also take much time for sentence analysis. In the present system there are 27 cluster symbols. They are shown in the Table with some examples. A simple explanation is given in the following lines.



N1: This symbol, though the same with a single noun, represents the noun phrase which has the possibility of being modified from both sides, such as AI (adjective) + N1 (noun) or N1 + N1.

N4: This symbol, which is the same with the pronoun, represents the noun phrase which can not be modified from the left side, for example, DT + N1 or DT + PT + N1.

PA: This represents the prepositional phrase which is mainly led by P1 class and takes some particles when it modifies noun or verb.

PB: This prepositional phrase mainly led by P2 class does not need any particles when it modifies noun or verb, the typical structure for this is P2 + N4.

L1: This phrase has such structure as TO + N4 , this is almost same with PA.

L2: This means "to-infinitive", and its typical structure is TO + VE. This symbol is translated into Japanese in many ways, as

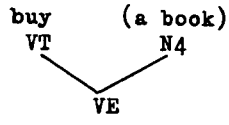
shown below,

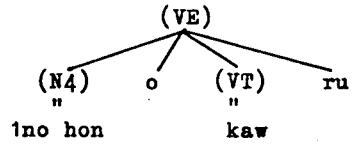
L2 + VE ⟶ SS	(L2) koto wa (VE)
N3 + VE + L2 ⟶ SS	(L2) koto wa (VE)
SS + L2 ⟶ SS	(L2) tameni (SS)
L2 + SS ⟶ SS	(L2) tameni (SS)
etc.	

RS: This is a relative clause which is led by relative pronoun R1 (who, which), R2 (whose), R3 (whom), and TH (that), and its structures are R1 + N4 + VE or R1 + VE , etc. This clause works as an adjectival clause or noun clause depending on the context.

WS: This has the same structure as RS if WN or WH replaces R1 or TH in RS.

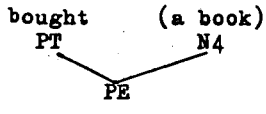
VE, GE, PE: As for verb phrases, there are three kinds of phrases VE, GE and PE according to their verb forms, that is, present, ing-form, and past form. The verb in phrase VE is already given a conjugation suffix, as below. On the other hand, the verbs in PE and GE are not

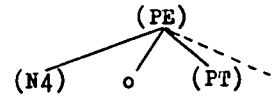




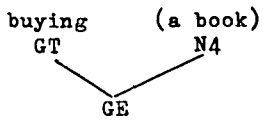
(1no hon o kawu)

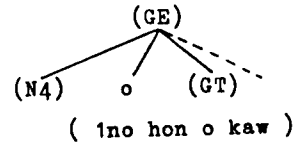
yet given conjugation suffixes, because PE and GE play many roles, such as adjectival or adverbial or predicative ones, and according to them the conjugation suffixes differ; therefore, the suffixes can not be determined until their roles are determined by the context. Their typical forms are as follows.



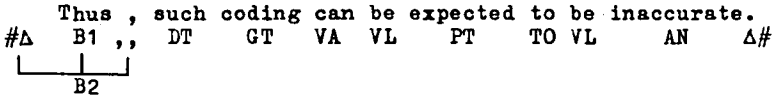


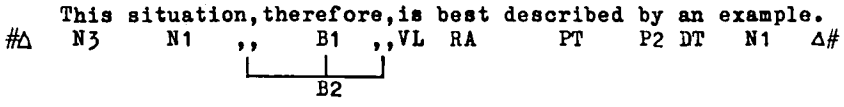
(1no hon o kaw)





B2: This symbol represents adverbial phrase or adverbial clause. In the next sentences some examples are shown.

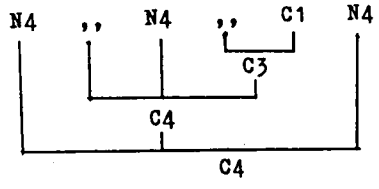




SS: This represents sentence form. The typical structure is N4 + VE.

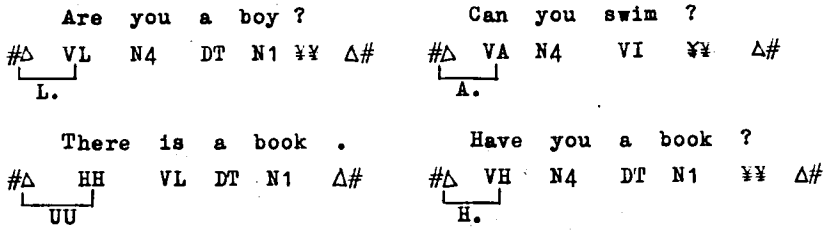
The above mentioned symbols represent the grammatically acceptable phrases and clauses. But there are some symbols which are used as conventional ones.

C3 and C4: These symbols are used to parse a marshaling of same kind words, for example, "N4, N4, and N4" or "VE, VE, and VE" etc. The structure of C3 and C4 are shown below.



Other symbols are those which characterize the words at the beginning of a sentence. If interrogative sentences need not be translated,

most of these symbols are unnecessary. Some of them are shown below.



In short, in this section, the fundamental attitude of word classification is described. They are as follows. (1) A special symbol is given to such words which cause the transition of word order, or require particles to be inserted in Japanese. (2) The name of phrase or clause must represent our grammatical concept as well as possible. (3) The number of symbols is desirable to be as small as possible.

3.5 Rewriting rules

Rewriting rules or patterns are rules for mechanical translation which are make of concatenation of several symbols. The general form is as follows.

$$\alpha \cdot \beta \cdot \gamma \longrightarrow \delta\ (\sigma\ ,\ \xi)$$

The left side symbols of arrow, $\alpha\beta\gamma$, consist of part of speech and represent some phrase or clause, and δ is a symbol which substitutes $\alpha\beta\gamma$. σ is a instructor for word order which indicates how the translation of α, β, γ is aranged in Japanese. ξ represents three symbols J_1, J_2, J_3 which are symbols for JYOSI which must be inserted in Japanese between the translation of α, β, and γ. Schematic representa-

tion of rewriting rule is shown below, where (x) means Japanese correspond to x.

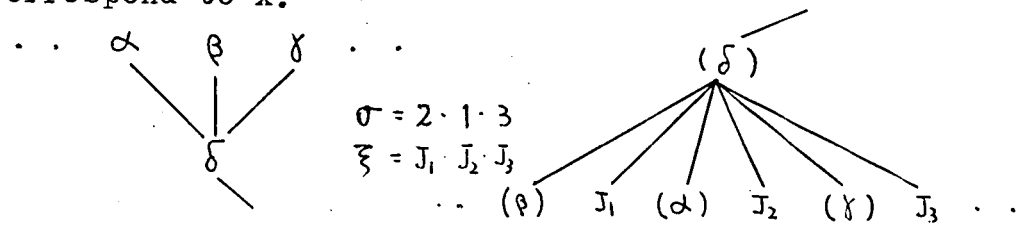


Fig. 3.5.1 Tree expression of rewriting rule $\alpha\cdot\beta\cdot\gamma \rightarrow \delta(2\cdot1\cdot3, J_1\cdot J_2\cdot J_3)$

The instructor σ = 2.1.3 means that the translation of the second word (β) in the sequence is arranged in the first position in Japanese, and the translation of the first one (α) is put in the second position (σ = 2 1.3), and the third one (γ) in English is put in the third position in Japanese. Then JYOSI, J_1, J_2, and J_3 are inserted in that order between translations. Another example is shown below. The instructor σ = 2·0·0 means that only second word in $\alpha\beta\gamma$, that is, β

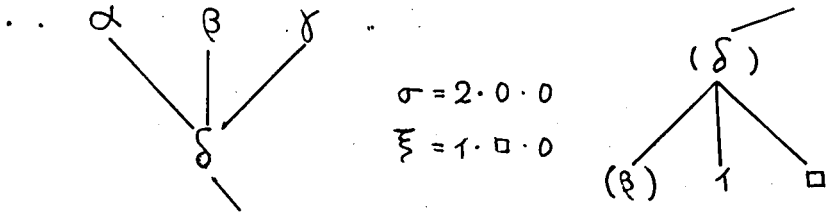


Fig. 3.5.2 Tree expression of rewriting rule $\alpha\cdot\beta\cdot\gamma \rightarrow \delta(2\cdot\phi\cdot\phi, 1\cdot\square\cdot\phi)$

is taken into consideration in Japanese. Then JYOSI 1 and □ are put after (β) in that order.

It is generally said that syntactic analysis in English can be done by dichotomy, that is, by the rule having two symbols on the left side of the arrow, $\alpha\cdot\beta \rightarrow \gamma(\sigma, \xi)$. Though dichotomy analysis may be

enough to analyze only English syntax, there are some cases where it can not exchange correctly the word order or insert appropriate JYOSI in translated Japanese. For example, three rewriting rules N4·PT ⟶ SS (1·2·0 , WA·IMASITA·Ø) can give a correct syntactic analysis for the part of sentence "...book which he bought..." as shown in Fig.3.5.3.

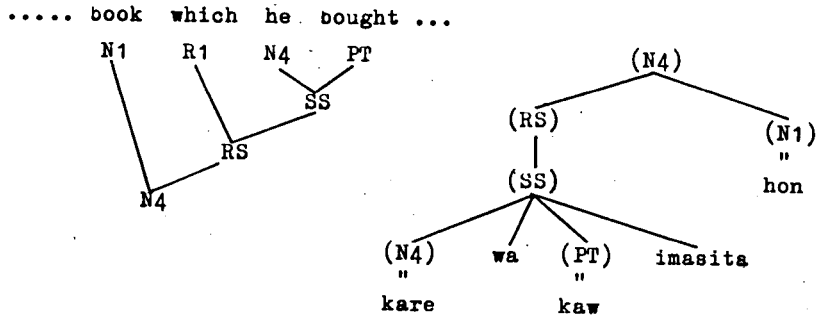


Fig. 3.5.3 An example which shows a dichotomy-analysis is not adequate for insertion of Japanese JYOSI.

But its translation becomes "...KARE WA KAWIMASITA HON...". The JYOSI, WA, is not correct in this case. This is because the fact that the subject in a subordinate clause requires usually JYOSI GA can not be considered in the dichotomy pattern. If a rule WI N4 PT ⟶ WS (2·3·Ø , GA·IMASITA·Ø) is added to the previous three rules, and given a higher hierarchy because of its wider context, then the above example becomes like Fig.3.5.4. Of course it is possible to get a correct result by using dichotomy rules which have a context-sensitive form, for example,

$$R1 \quad \overset{1}{N4} \quad \overset{2}{PT} \longrightarrow R1 \quad SS \quad (1\cdot 2 \ , \ ga\cdot imasita)$$

but context-sensitive rules are tedious to apply mechanically. If only they are context-free form, two symbol or three symbol pattern are processed with quite the same procedure.

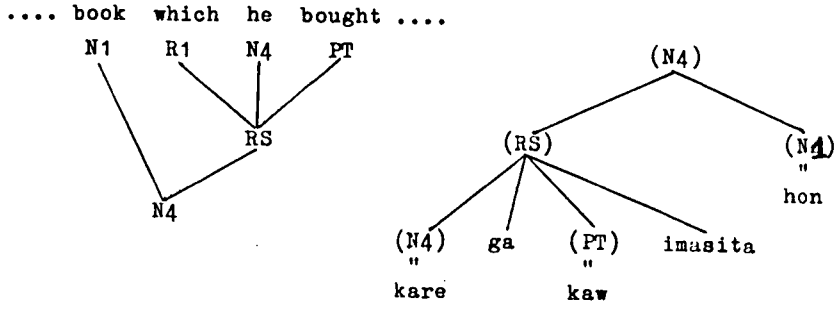


Fig. 3.5.4 An adequate analysis for insertion of a correct Japanese JYOSI /ga/.

In the above example, three symbol patterns are used only to make translated Japanese more natural, and not because it is essentially required in English syntax analysis. Are there three term relations really in languages? In English the next example (Fig.3.5.5) may be one of the rare cases.

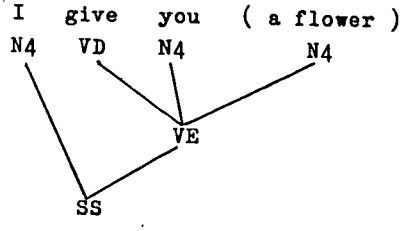
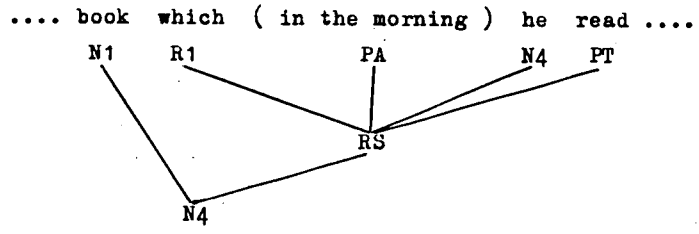


Fig. 3.5.5 An example of essentially three term relation.

As for the four term relations in essence, there are perhaps no such relations in natural languages or even in mathematics. The next examples, however, can be looked as a four term relation if a simple correspondence between English and Japanese is taken into consideration, though it may be for the sake of convenience, and not essencial.



In this example, a prepositional phrase PA is inserted irregularly, so context-free rules requires long sequence to predict correct JYOSI /GA/ for the subject of a subordinate clause. Except these irregular cases, there may be scarcely four term relation. Therefore in this translation system two term and three term patterns are thought to be enough to analyze and synthesize large parts of English and Japanese.

It must be noted that, in the present system, a once parsed string can not be separated or disjoined again. For example, suppose that a sequence "...a b c..." becomes "...y..." and its word order in Japanese is (b)(a)(c), by the rules $a \cdot b \longrightarrow x$ $(2 \cdot 1 , \emptyset \cdot \emptyset)$ and $x \cdot c \longrightarrow y$ $(1 \cdot 2 , \emptyset \cdot \emptyset)$.

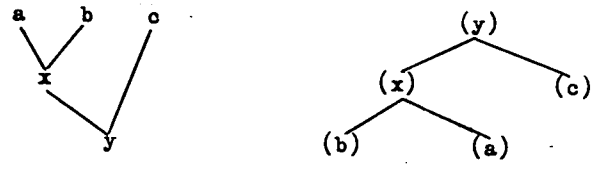
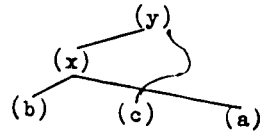
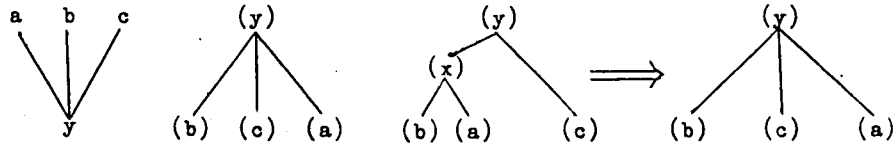


Fig. 3.5.6 Two-term context-free correspondence.

If it is a rule in the translated language that (c) always precede to (a), then the structure of translated tree must be changed as below.

This procedure, however, is very bothersome, though flexible as a system, so these changes are not allowed in the present system. This can be treated by mapping, or by the rewriting rule $a \cdot b \cdot c \longrightarrow y\ (2 \cdot 3 \cdot 1\ ,\ \emptyset \cdot \emptyset \cdot \emptyset)$.

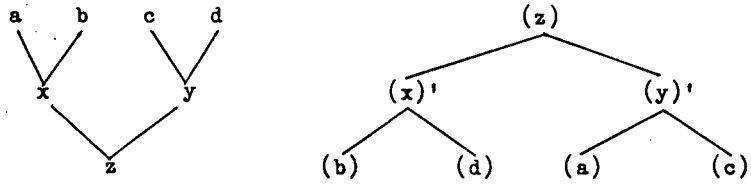




Three-term rule Mapping

Fig. 3.5.7 Correct tree bythree-term rule or mapping.

Therefore even fixed context-free rules are very useful. But if such a case as is shown below happens, rewriting rules which have only three symbols on the left side of the arrow can not treat.



English Japanese

In the following, various kinds of patterns are explained using concrete examples.

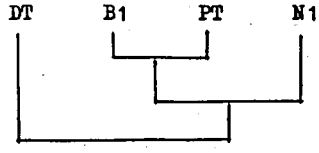
3.5.1 Noun phrase

Most frequent patterns are noun phrases which are modified by determiners, adjectives or their equivalent words from the left side

of the main noun. Some typical cases are shown below.

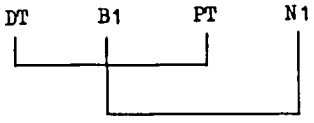
1	a book	DT · N1 → N4 (1·2 , ø·ø)
2	many books	AO · N1 → N1 (1·2 , no·ø)
3	a beautiful flower	DT · AI · N1 → N4 (1·2·3 , ø·I·ø)
4	cat's eyes	N1 · '' · N1 → N1 (1·3·ø , no·ø·ø)
5	punched - card	PT · -- · N1 → N1 (1·2·3 , rareta·ø·ø)
6	the dotted line	DT · PT · N1 → N4 (1·2·3 , ø·rareta·ø)
7	a starting point	DT · GI · N1 → N4 (1·2·3 , ø·itutuary·ø)
8	a well formed formula	DT · B1 · PT → DT (1·2·3 , ø·ø·rareta)
		DT · N1 → N4 (1·2 , ø·ø)

There are many other examples which are transfigurations of the above examples, for example, "a pretty little girl school" is analyzed by the combination of 1 and 2. A phrase "an___ly___ed___" is the same structure as DT PV N1 ,if a pattern B1·PV → PV is applied first. But as mentioned in section 3.2, adverbs which modify verb phrase from the left side of the verb is given a lower hierarchy than right side modifier. So itcan not be applied at first. In this case, however, the sequence of part of speech "...DT B1 PT N1..." determine that the word named PT is used as an adjectival word, so such a rule DT·B1·PT → DT (1·2·3 , ø·ø·RARETA) is introduced, though it does not obey the modification rule mentioned in section 3.2.



True structure

Fig. 3.5.10 Two structures of DT B1 PT N1 (to be continued).



Conventional structure

Fig. 3.5.10 Two structures of DT B1 PT N1. (continued)

In the above example a determiner "DT" plays a leading role in determining the function of PT, so if the determiner is missed, it is difficult to predict that PT works as adjectival.

There are several rules which have almost the same structure each other, such as DT·PI·N1 ⟶ N4 (1·2·3 , ∅·IMASITA·∅) DT·PT·N1 ⟶ N4 (1·2·3 , ∅·IMASITA·∅), and DT·PD·N1 ⟶ N4 (1·2·3 , ∅·IMASITA·∅), they differ only in the second symbol. PI PT and PD are sub-classes of verbs, so these rules can be merged into one form DT·PV·N1 ⟶ N4 (1·2·3 , ∅·IMASITA·∅). But the algorithm to compare pattern with the sequence in the given sentence becomes rather tedious. If the above three rules are stored in the pattern table, the comparison algorithm becomes a simple table-look-up, though the number of rules increases.

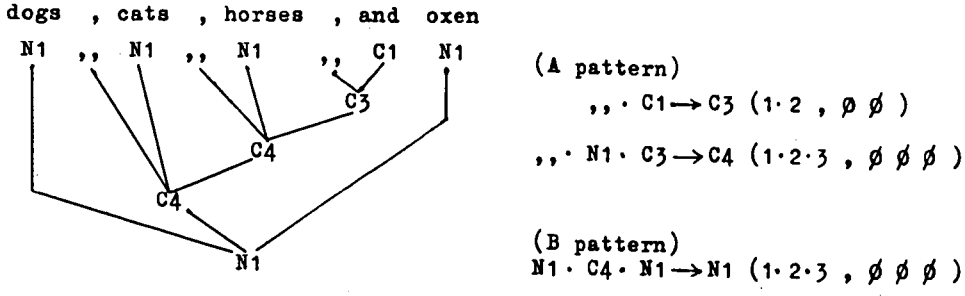


Fig. 3.5.11 Parsing of itemized nouns by context-free rules.

Frequently several nouns are arranged in a low as shown in Fig. 3.5.11. Since it is not so easy to parse this sequence into noun phrase by context-free rules, somewhat convenient symbols C3 and C4 are introduced. These rules are only effective when the input sequence has such a standard form as shown in Fig.3.5.11. If the input sentences are "A, B, C and D" or "A, B, C, D", then they must be changed to a standard form by simple pre-editing.

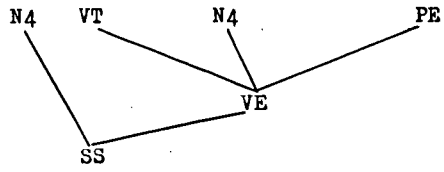
The noun phrases in which the main nouns are modified by several phrases or clauses from the right side are shown below.

1	(a flower	)(in the vase)		
	N4	PA		N4·PA→N4 (2·1 , ni(no)·ø)
2	(a flower	)(which she gave me)		
	N4	RS		N4·RS→N4 (2·1 , ø·ø)
3	see (an example	)(mentioned in this paper)		
VT	N4	PE		VT·N4·PE→VE (3.2 1 , rareta·o·ru)
4	(an example	)(showing a noun phrase)		
	N4	GE		N4·GE→N4 (2·1 , itutuaru·ø)
5	use (all data	)(available)		
VT	N4	AN		N4·AN→N4 (2·1 , na·ø)

Fig. 3.5.12 Noun phrases in which the main nouns are modified from the right side to them.

As seen from Fig.3.5.12, the right side modifiers for nouns are prepositional phrase (PA, PB), relative clause (RS), verb phrase led by the past form verb (PE) or ing-form verb (GE), and some special adjectives (AN). As for the detailed explanation of PA PB RS PE and GE, it will be given later in this section. The rewriting rules in Fig.3.5.12-(3) except VT·N4·PE ⟶ VE may be easily understood, but VT·N4·PE ⟶ VE is a complex one. The verb phrase in past tense does not necessarily work as an adjectival phrase only because it is situated just behind a noun. It depends on a larger context. A typical example is shown below (Fig.3.5.13).

(1) He wants (a result) (obtained by the computer).



(2) He (obtained by the computer).

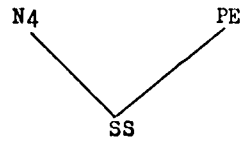
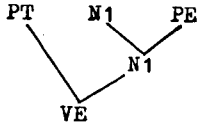


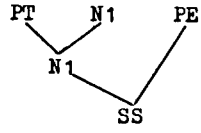
Fig. 3.5.13 Different function of PE according to the different contexts.

In the example (1), PE is recognized as a modifier by the context "...VT N4 PE...", and the rewriting rule VT·N4·PE ⟶ VE (3·2·1 , RARETA·O·RU) leads to the correct translation, "カレワ ソノ コンピュータアニヨッテ テニイレラレタ イノケッカヲ ホッスル". The reason why PE is recognized as a modifier is that in the scientific papers the tense of main verbs is usually present, so if there exist present form and past form verbs in the near position as in the above case, it is natural to infer that VT is a verb and PE led by a past form verb works as an adjectival phrase. On the other hand, in the example (2), the same phrase PE is looked as a main predicate and it is translated into "カレワ ソノ コンピュータアニヨッテ テニイレマシタ " by the rule N4·PE ⟶ SS (1·2 , WA·IMASITA). The only difference between (1) and (2) is that in (2) there is no present form verb just before the noun phrase N4. In the next example (3), a verb just before

(3) He wanted results (obtained by the computer).
N4 PT N1 PE

the noun N1 has a past tense, then there are two possibilities in syntactic structure, as below.





The former case is the same structure as (1), but in the latter case PT is looked as an adjectival use as in the case "...stored program..", and PE as a main predicate. It must be noted that it is not N4, but N1 which appears between PT and PE in this case. If N4 appears in the position of N1, the structure PT(N4 PE) is possible. These ambiguities become definite by looking a larger context, but for this purpose, re-writing rules must contain long sequence of parts of speech, but they become more tedious to be applied. Therefore at the risk of making several errors, the rule PT·N1·PE $\longrightarrow$ PE (3·2·1 , RARETA·Ø·RU) is adopted from the probabilistic point of view.

As for GE, it is not so ambiguous as in the case of PE, and almost always GE just behind N4 or N1 can be thought to be a modifier, and its rule is N4·GE $\longrightarrow$ N4 (2·1 , TUTUARU·Ø). A concrete example is shown below.

" It is gratifying to receive news that
the resolusion calling for an immediate ceasefire
N4 GE
was passed unanimously tuesday night, New York time"

It is possible to apply just the same rule to the case where an ing-verb phrase is thought to be complementary in ordinal grammar, as shown

below.

I saw (the thief) (running away).
N4 PT N4 GE

N4 GE → N4 (2·1 , itutuaru·ø)
watasi wa nigetutuaru doroboo o mita
PT N4 GE → PE (2·3·1 , ga·ru no o·ø)
watasi wa doroboo ga nigeru no o mita

If such a rule as PT·N4·GE ⟶ PE (2·3·1 , GA·RU·NO·ø) is used, then the translation of above example become more natural in Japanese as "ワタシ ワ ドロボウ ガ ニゲル ノヲミタ". But the adequacy of this rule, especially in JYOSI, depends on the semantic aspect of verb VT.

It is difficult, however, to treat a participial construction which has two different subjects. For example,

" The sun having set, we put up at inn."
N4 GE ,, N4 PE

in this case GE can not be recognized correctly by simple pattern as a predicate. If the subject of a participial construction is the same as that of the main clause, it is possible to judge that GE works as a subordinate clause by the rules #Δ·GE,, ⟶ B2 (2·3 , ITUTU·ø) or SS·,,·GE → SS (3·2·1 , ITUTU·ø·ø) , ,,·GE·Δ# → B2 (2·1·ø , ITUTU·ø·ø).

Using the Multilist technique , the total information has been integrated.

#Δ GE ,, SS Δ#
B2

One possibility is the preplanning , preparing control cards in advance.

#Δ SS ,, GE Δ#
B2

Fig. 3.5.14 GE working as a subordinate clause.

Some adjectives constitute similar structures to those of GE. They modify nouns just before them,

(the example) available in this case
N4 AN

(a difference) appreciable to the eye
N4 AN

the sun keep (us) warm
N4 AI

In the third example, the word "warm" is used as a complement, and it is more natural to treat it by the rule VT·N4·AI ⟶ VE (2·3·1 , Ø·KURU). As for the first example, N4·AN ⟶ N4 (2·1 , NA·Ø) translate "The example available in this case " into "KONO BAAI NI RIYOKANONA REI". This two-term rule, however, may give erroneous results to interrogative sentences. It is also the same as the case of GE. For example,

Is (her sister) (working at the desk) ?
VL N4 GE ¥¥

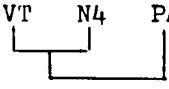
Is (this milk) (warm) ?
VL N4 AI ¥¥

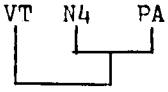
These cases must be translated by VL·N4·GE ⟶ SS (2·3·1 , WA·ITUTU·RU) and VL·N4·AI ⟶ SS (2 3 Ø , WA·I·Ø).

As for the prepositional phrase which is situated behind a noun phrase, there are ambiguous cases from semantical point of view. That is, such a prepositional phrase does not necessarily modify the noun just before it, but may work as an adverbial phrase. It is impossible to determine which is which by the syntactic information. According to the correspondence of word order in Japanese with that of English, which is mentioned in section 3.2, the word order in Japanese is the

same in both cases, though particles which accompany the prepositional phrase in Japanese differ in the case of an adverbial use from an adjectival use. A few examples are shown below.

He built a house on the hill. (adverbial)	kare wa oka no ue ni ie o tateta.
He lives in a house on the hill. (adjectival)	kare wa oka no ue no ie ni sumu.

It will be one method to use an ambiguous JYOSI NI(NO), thinking that the prepositional phrase always modifies the noun when it comes after the noun. Therefore instead of a correct structure 

, an apparently a wrong analysis 

is adopted. But in Japanese both cases are correct in word order, "PA に(の) N4 と VT" and "PA に(の) N4 と VT".

3.5.2 Prepositional Phrase

The general structure of a prepositional phrase consists of a preposition followed by the noun phrase mentioned in 3.5.1, or noun equivalents, such as Pl·N4 ⟶ PA (2·1 , Ø·Ø) or Pl·GE ⟶ PA (2·1 , RUKOTO·Ø), etc. A few examples which belong to rather a special case are shown.

in variably formatted pages (N1)	P1 B1 N1 → PA (2·3·1 , Ø·Ø·Ø)
in operating speed	P1 GT N1 → PA (2·3·1, ITUTUARU·Ø·Ø)
with user-specified-labels (N1)	P2 N1 → PB (2·1 , Ø·Ø)
by adding the appropriate plot (GE)	P2 GE → PB (2·1 , RUKOTO·Ø)
(routines) for common analysis (N1)	P4 N1 → PB (2·1 , Ø·Ø)

In the second example, it is difficult to infer whether the word "operating" works as an adjectival or gerund, and it depends on the semantic aspect of the verb and noun. From the syntactic point of view, it can be only said that an ing-form and a past-form verb are probably used as an adjectival word, if they are situated between a preposition and a single noun. Prepositions play the same role as determiners in this respect. Then the rules have the next form, P1·GI·N1 ⟶ PA (2·3·1 , ITUTUARU·Ø·Ø) or P1·PI·N1 ⟶ PA (2·3·1 , RARETA·Ø·Ø), etc.

3.5.3 Verb phrases

Typical forms of verb phrase are as follows.

It is easy. (VL AI)	VL · AI ⟶ VE (2·Ø , i·Ø)
I am a boy. (VL N4)	VL · N4 ⟶ VE (2·1 , de·ru)
She follows him. (VI N4)	VI · N4 ⟶ VE (2·1·, ni·ru)
I give him a book. (VD N4 N4)	VD N4 · N4 ⟶ VE (2·3·1 , ni·o·ru)

Attention must be paid to the conjugation in Japanese with respect to these rules. The translation word for an English verb is given in the

dictionary in its stem form of present tense, therefore the conjugation in Japanese must be taken into consideration when a rule is applied.

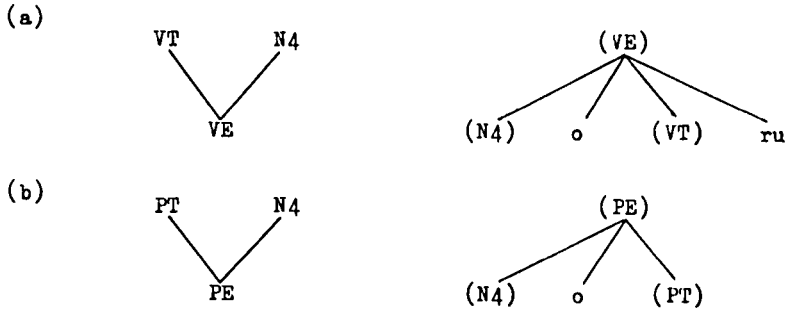


Fig. 3.5.19 Structure of VE and PE.

As a principle, conjugation particles such as RU, IMASITA, RARETA, etc., are inserted when the role of a verb is determined by the context. For example in Fig.3.5.19, the present tense verb is clearly understood to be a predicate, so that conjugation particles for the present tense "RU" is given. But the verb phrase led by the past form verb in the second example can not have its role determined by such a sort of string as PT N4, and then any conjugation particles are not given, though they are substituted by PE. The conjugation of PE is determined, for example, by the next context (Fig.3.5.20), the branch enclosed by the dotted line

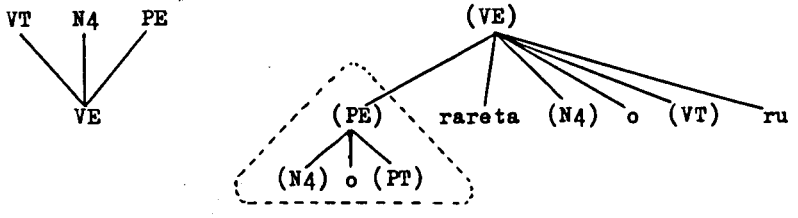


Fig. 3.5.20

shows the previously constructed tree in the second example of Fig. 3.5.19. It is also the same as the case of ing-form verb. An example

is shown with a complete tree structure.

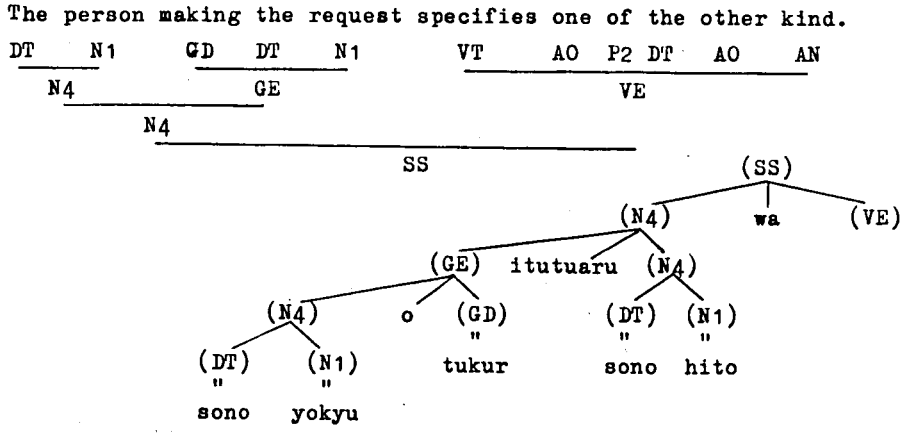


Fig. 3.5.21 Correspondence tree between English and Japanese.

In the case of the dative verb (VD), the rewriting rule VD N4 N4 $\longrightarrow$ VE (2·3·1 , N1·O·RU) is applied. Several other verbs concerning perception, causality, etc. require verb phrases as complements. For example

I found him working at his desk.

I like to hear her sing.

I must get my hair cut.

I will help you carry that box upstairs.

In the above examples, the main verbs are all named VT (transitive verb), but if they are classified into sub-classes according to their meaning and usage, then their translations become more Japanese-like, and also less ambiguous in syntax.

3.5.4 Tense and Mood of verb

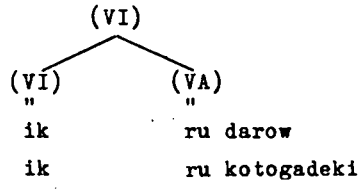
The tense of verb is also processed by using a fixed length context-free rewriting rules. In table 3.5.4.1, all cases of tense are listed. Some of them are explained below.

Future tense and Potential mood: An affirmative case of future tense or potential mood is treated by the next rule. Auxiliary verbs (will, must can etc.) are included in the same class VA, so future tense

VA VI ⟶ VI (2 1 , ø ø)

will go

can go



is not distinguished from potential mood. This leads to a inadequate expression for a negative case of future tense. Japanese words which correspond auxiliary verbs in English have somewhat queer forms; will = *U DAROW , can = *U KOTOGADEKI , etc. But these expressions are inadequate for interrogative sentenses, so it is desirable to change their form as follows; will ; DAROW , can ; KOTOGADEKI , etc. According to these change, the rewriting rule also must be changed as below.

VA VI ⟶ VI (2 1 , ru ø)

Perfect tense: In Japanese, roughly speaking, there is no such tenses corresponding to the English perfect tense, it is included in the past tense. Therefore the perfect tense in English is substituted by past tense in this translation system. That is, "have studied" and "had studied" are equivalent to "studied". As for the negative case of perfect tense, it is somewhat difficult to insert an adequate particles,

Table 3.5.4.1 Tense table

	Present	Past	Future
Infinit	VT	PT	VA VT
	KAK-u	KAK-imasita	KAK-u darow-u
Negative	VF EE VT	PF EE VT	VA EE VT
	KAK-an-u	KAK-an-akatta des-u	KAK-an-u darow-u
Passive	VL PT	PL PT	VA VL PT
	KAK-are-ru	KAK-are-masita	KAK-are-ru darow-u
Negative	VL EE PT	PL EE PT	VA EE VL PT
	KAK-are-n-u	KAK-are-nakatta des-u	KAK-are-n-u darow-u
Perfect	VH PT	PH PT	VA VH PT
	KAK-imasita	KAK-imasita	KAK-imasita darow-u
Negative	VH EE PT	PH EE PT	VA EE VH PT
	KAK-anakatta des-u	KAK-anakatta des-u	KAK-anakattades-u darow-u
Passive	PH PL PT	PH PL PT	VA VH PL PT
	KAK-are-masita	KAK-are-masita	KAK-are-masita darow-u
Negative	VH EE PL PT	PH EE PL PT	VA EE VH PL PT
	KAK-are-nakatta des-u	KAK-are-nakatta des-u	KAK-are-nakatta des-u darow-u
Progressive	VL GT	PL GT	VA VL GT
	KAK-itutuar-u	KAK-itutuar-imasita	KAK-itutuar-u darow-u
Negative	VL EE GT	PL EE GT	VA EE VL GT
	KAK-itutuar-an-u	KAK-itutuar-anakattades-u	KAK-itutuar-an-u darow-u
Perfect Progressive	VH PL GT	PH PL GT	VA VH PL GT
	KAK-itutuar-imasita	KAK-itutuar-imasita	KAK-itutuar-imasita darow-u
Negative	VH EE PL GT	PH EE PL GT	VA EE VH PL GT
	KAK-itutuar-anakattades-u	KAK-itutuar-anakattades-u	KAK-itutuar-anakatades-u darow-u

have VH	studied PT	VH · PT ⟶ PT (2·ø , ø·ø)
had PH	studied PT	PH · PT ⟶ PT (2·ø , ø·ø)

Fig. 3.5.22 Perfect tense.

sa such a rule as VH·EE·PT ⟶ VT (3·ø·ø , ANAKATTADES·ø·ø) is used. For example

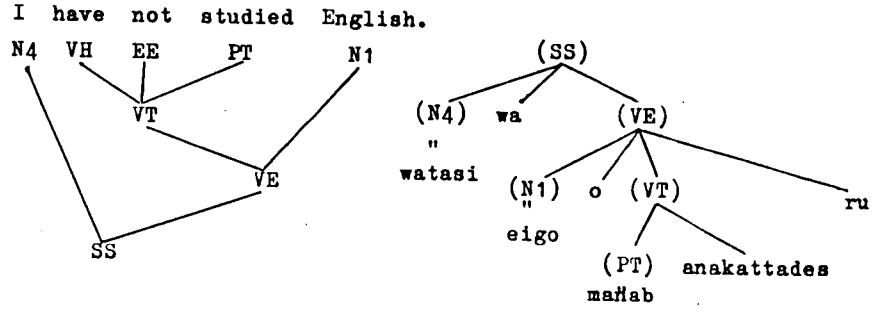


Fig. 3.5.23 Negative case of perfect tense.

Passive: Passive form is processed by next rule.

VL · PT ⟶ VI (2·ø , rare ø)

PL · PT ⟶ PI (2·ø , rare ø)

In this rule, a substituted symbol is VI, not VT, because the object of PT is placed before VL (subjective case), and VL PT works as if it is intransitive verb. Therefore if noun or objective are there just behind PT, then the word will be complement. In the case of dative verbs, it is substituted by transitive verb, as below.

VL · PD ⟶ VT (2·ø , rare ø)

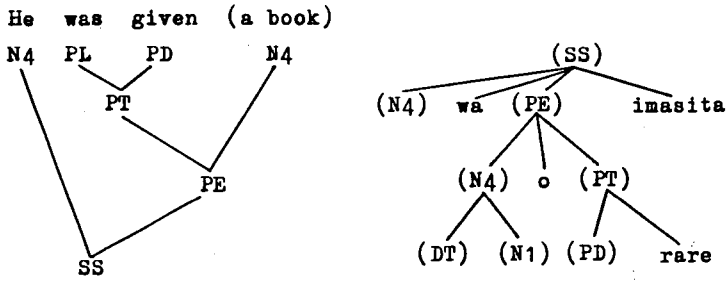


Fig. 3.5.24 Passive form of dative verb.

Progress tense: Simple progress tense is processed by the next rule,

VL GT —— VT (2 1 , itutu ∅)

PL GT —— PT (2 1 , itutu ∅)

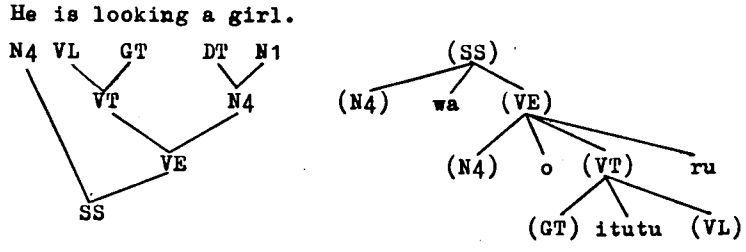


Fig. 3.5.25 Progress tense.

As for present perfect progress, combinations of progressive tense and perfect tence are used.

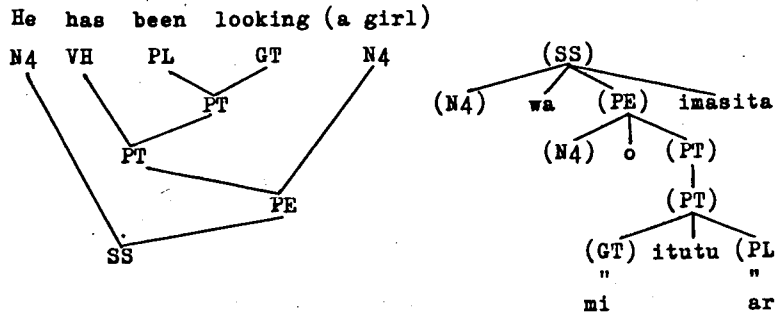


Fig. 3.5.26 Present perfect progress form.

By the way, the treatment of negative case will be shown below. The most simple types are as follows.

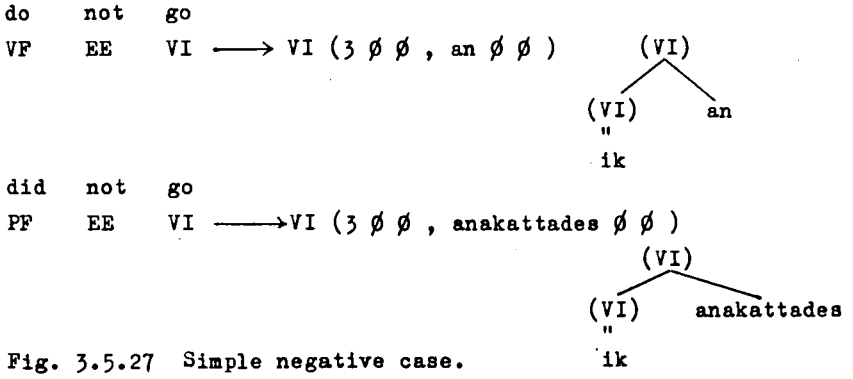


Fig. 3.5.27 Simple negative case.

In the second example, a rather crude conjugation suffix (-ANAKATTADES) is inserted, and a substitute word is VI not PI. The reason is as follows. The negative conjugation suffix of past tense in Japanese is " -ANAKATTA" or " -IMASENDESITA". These suffixes can not be constructed by the combination of the negative particle "-AN" and past particles "-IMASITA". In other words, it belongs to the irregular expression. Then to make it apparently regular such crude suffixes are introduced. It is possible, however, to conjugate in a more natural way if new parts of speech are introduced. That is, if a new symbol PI' is introduced and, instead of the second rule in above example, new rules PF·EE·VI ⟶ PI' (3·ø·ø , ANAKATTA·ø·ø), N4·PI' ⟶ SS (1·2 , WA·ø), etc., which are obtained by substituting PI by PI' and erasing JYOSI "IMASITA", are introduced, the negative case of past tense can be naturally translated into Japanese as shown in the next page.

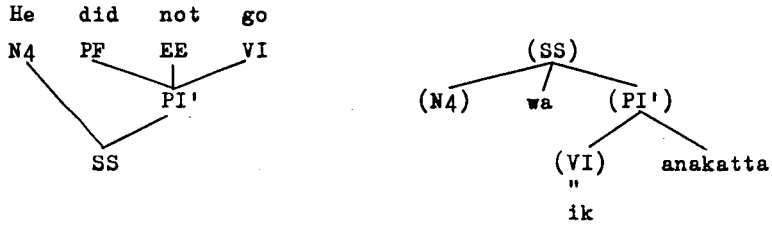


Fig. 3.5.28 Adequate structure of negative case of past form.

The introduction of new symbols increases the number of rules, then, it is desirable, if possible, to make shift with existing symbols without any confusion. If PF EE VI is substituted by PI, it can not be distinguished from the single past form word, and past tense conjugation suffix for affirmative case "-IMASITA" is given. To avoid this, PF EE VI is substituted by VI, and suffix "-ANAKATTADES" is inserted to match the present form suffix "RU" when it is given to VI. An example is given below. The translated sentence becomes "KARE WA IKANAKATTADESU".

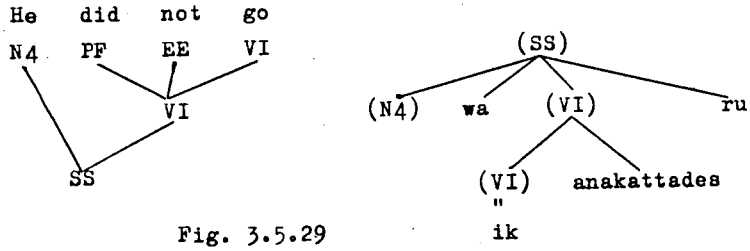


Fig. 3.5.29

The combination rule for conjugation suffixes are given in later section.

Another examples of negative case are shown below only in tree structure.

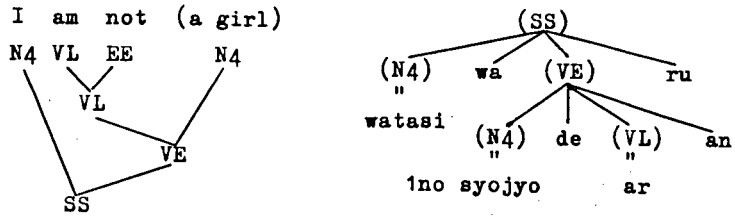


Fig. 3.5.30 An example of negative case.

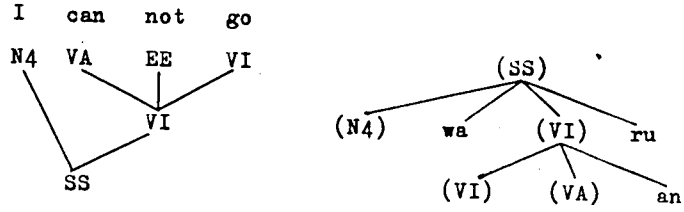


Fig. 3.5.31 An example of negative case.

3.5.5 Relative clause

Typical form of relative clauses, which are led by relative pronouns or relative adverbs, are shown below. Usually the translation word of

R1 · N4 · VT ⟶ RS (2·3·Ø , ga·ru·Ø)

W1 · N4 · VT ⟶ WS (1·2·3 , Ø·ga·ru)

R2 · N1 · VE ⟶ RS (1·2·3 , Ø·ga·Ø)

R1 · VE ⟶ RS (2·Ø , Ø·Ø)

C2 · N4 · VE ⟶ B2 (2 3·1 , ga·Ø·Ø)

Fig. 3.5.32 Some examples of relative clauses and adverbial clause (the last one).

a relative pronoun itself is neglected in Japanese. In a special case the translation word appears, as below.

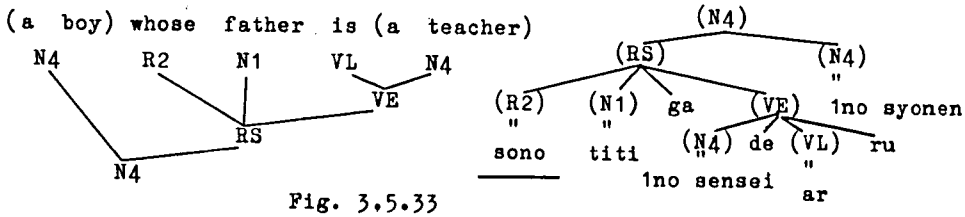
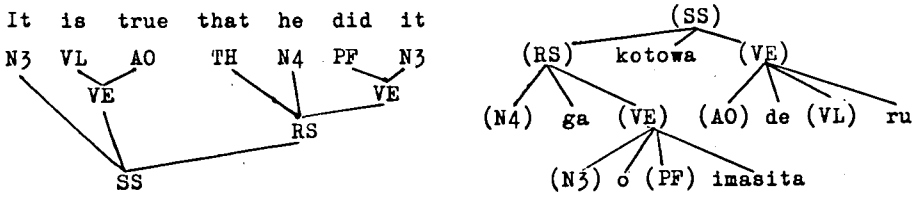


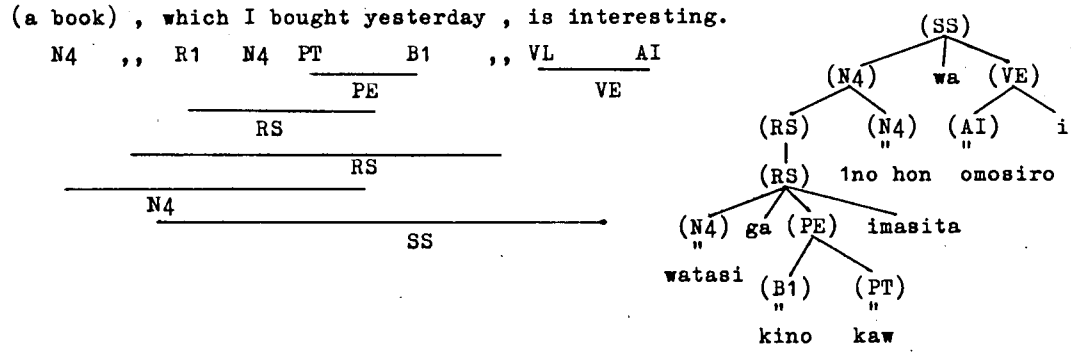
Fig. 3.5.33

Other examples which include relative clause are shown.



(kare ga sore o simasita kotowa honto de aru)

Fig. 3.5.34 An example of "it-that-clause".



(watasi ga kino kawimasita ino hon wa omosiroi)

Fig. 3.5.35 An example of an inserted relative clause.

In this example a relative clause is inserted. These cases may be thought to be difficult to analyze in a ordinary method, but it is easily recognized by context-free rules that the relative clause modify noun-phrase "a book".

3.5.6 Sentence

The most typical form of sentence is like this,

$$\underset{N4}{I}\ \underset{VE}{\text{am a boy.}} \qquad N4 \cdot VE \longrightarrow SS\ (1 \cdot 2\ ,\ wa \cdot \emptyset\ )$$

$$\underset{HH}{\text{There}}\ \underset{VL}{\text{is}}\ \underset{N4}{\text{a book.}} \qquad HH \cdot VL \cdot N4 \longrightarrow SS\ (3 \cdot 2 \cdot \emptyset\ ,\ ga \cdot ru : \emptyset\ )$$

Sometimes to-infinitives appear in the position of subject,

$$\underset{N3}{\text{It}}\ \underset{VE}{\text{is easy}}\ \underset{L2}{\text{to go.}} \qquad N3 \cdot VE \cdot L2 \longrightarrow SS\ (3.2\ \emptyset\ ,\ kotowa \cdot \emptyset \cdot \emptyset\ )$$

$$\underset{\#\Delta}{}\ \underset{L2}{\text{To know}}\ \underset{VE}{\text{is difficult.}} \qquad \#\Delta\ \ L2\ \ VE \longrightarrow SS\ (2 \cdot 3 \cdot \emptyset\ ,\ kotowa \cdot \emptyset \cdot \emptyset\ )$$

If N3 is replaced by N4 in the above rule, it becomes another structure

$$\underset{N4}{I}\ \underset{VE}{\text{eat foods}}\ \underset{L2}{\text{to live.}} \qquad N4 \cdot VE \cdot L2 \longrightarrow SS\ (1 \cdot 3 \cdot 2\ ,\ wa \cdot tameni \cdot \emptyset\ )$$

Interrogative sentences generally have such a structure as "do + SS + ? ", for example,

Do	you	want	the	book	?
VF	N4	VT	DT	N1	¥¥
		SS			

Then the next rules are used to translate it.

$$SS \cdot ¥¥ \longrightarrow SS\ (1 \cdot \emptyset\ ,\ ka \cdot \emptyset\ )$$

$$VF \cdot SS \longrightarrow SS\ (2 \cdot \emptyset\ ,\ \emptyset \cdot \emptyset\ ) .$$

In the simple interrogative sentence, "do(does)" can be thought to be redundant from the syntactic point of view, because it can be recog-

nized as an interrogative sentence by the existence of interrogative mark without do-verb. In the case of "Can you play the piano ?", such

$$VA \cdot SS \circ ¥¥ \longrightarrow SS\ (2\ 1\ \emptyset\ ,\ \emptyset\ ru\ ka\ )$$

a rule must be used. If $VA \cdot SS \longrightarrow SS\ (2 \cdot 1\ ,\ \emptyset \cdot RU)$ is used instead of the above rule, the result of the above example becomes like below because of right to left passing.

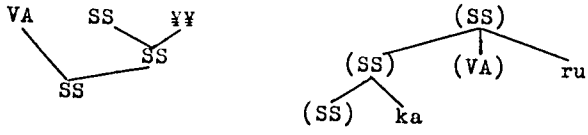


Fig. 3.5.36

But $VA \cdot SS \longrightarrow SS\ (2 \cdot 1\ ,\ \emptyset \cdot RU)$ is necessary in the case of "Can you go to the station, Jack ?". This analysis tree becomes as below.

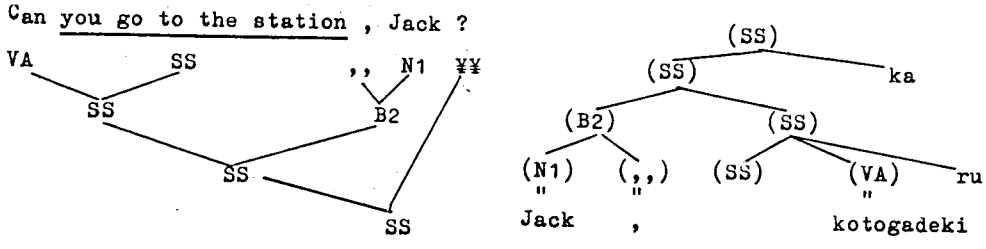
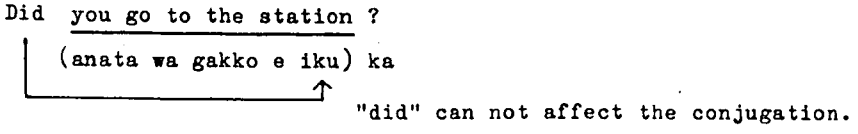
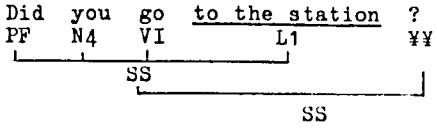


Fig. 3.5.37

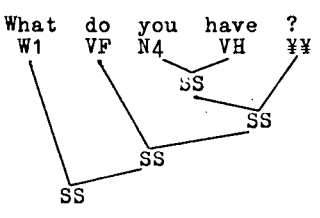
Though these cases seldom appear in scientific papers, they are shown to show that they can be treated by pattern method. As for the past form of interrogative sentence, however, it is difficult to give a correct conjugations, because the main verb in the bracket have already been conjugated and the past tense conjugation can not be given without disjoining connected words, which, however, is prohibited in this system.

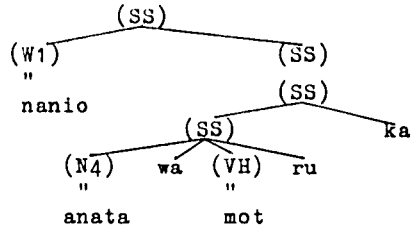


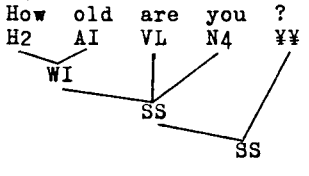
If the four-term rule is used, it is correctly processed, but the number of patterns which seldom appear increase.

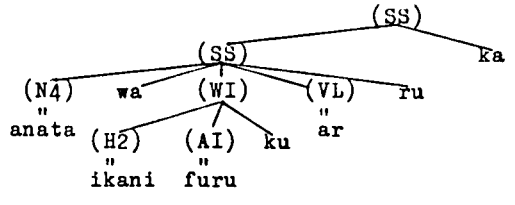


Other interrogative sentences are shown.



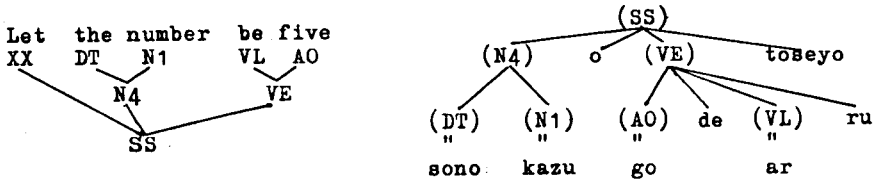




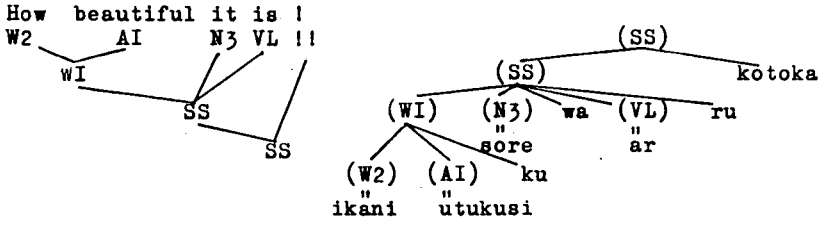


It is very difficult to treat such a sentence as "What time is it now ?" or "Who is that ?" by the pattern method only, they must be treated as idiomatic expressions. Imperative sentences are processed by the next rules.

Let's go for a walk.	XX · VE → SS (1 2 , ∅ besi)
Let the number be five.	XX · N4 · VE → SS (2 3 ∅ , O·toseyo·∅)
Open the window.	#Δ · VE → SS (2 ∅ , besi ∅)



Exclamatory sentences, if necessary, can be processed like below,



3.5.7 Other structures.

If sentences have a standard structure, though it is a question what is a standard, a large part of them can be treated by the rule-by-rule translation. Actual sentences surely have, in a local, standard phrase structures, nevertheless it is difficult to recognize as a whole how each part relates to other parts. For example,

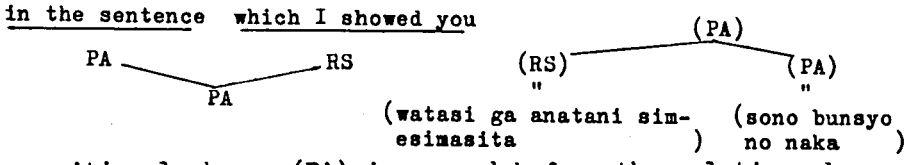
> " It is certainly desirable to have document index, that is, descriptor lists, as short as possible."

> " Such structure have been popularized in recent years in ,for example,pert networks, and lend themselves readily to automatic data processing methods."

they seem to be very natural and not complex, but it is not so easy to recognize by the patterns what role the underlined parts play in the whole sentence. As regards the above example, they may be solved by looking at such a phrase ", that is," or ",for example," as idiomatic expressions, and by giving them such parts of speech C1 (conjunction) or B2(adverbial phrase). But, generally speaking, the treatment of a inserted phrase or adverbial phrase is one of the difficult problems in pattern method.

Several rules, which seem to be rather adhoc, are explained in the following.

PA·RS ⟶ PA (2·1 , Ø·Ø) : In this rule, it seems as if the relative clause (RS) modifies the prepositional phrase (PA), but in the translated sentence, RS modifies the noun phrase which is included in the prepositional phrase. This case occurs frequently, for example, in the next sentence.



The prepositional phrase (PA) is parsed before the relative clause modifies the noun phrase "the sentence", because the rules which include the verb are lower in hierarchy than the noun phrase or prepositional phrase. In this case, however, the structure of PA is known to be "preposition + noun phrase", and the structure of its translation is also known to be "noun phrase + preposition" according to the rule P1·N4 ⟶ PA (2·1 , Ø·Ø); therefore, by the rule PA·RS ⟶ PA (2·1 , Ø·Ø), it is possible to make such word order as if RS modifies the noun phrase in PA. In Fig.3.5.4.2, the right side structure is correct, but the word order in both structures are the same.

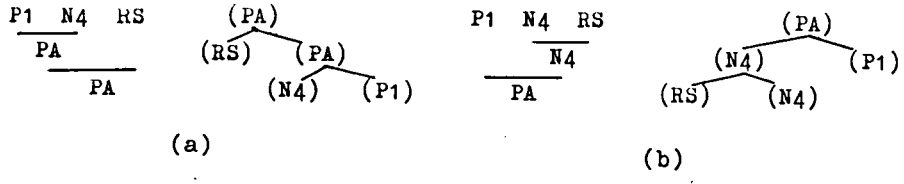


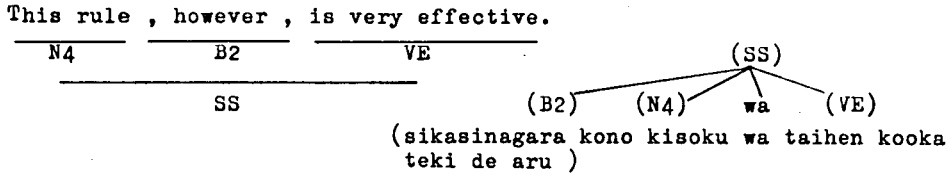
Fig. 3.5.4.2

There are several rules which are similar to this case.

in a system of searching files　　PA · PB → PA (2·1 , ∅·∅)
PA PB

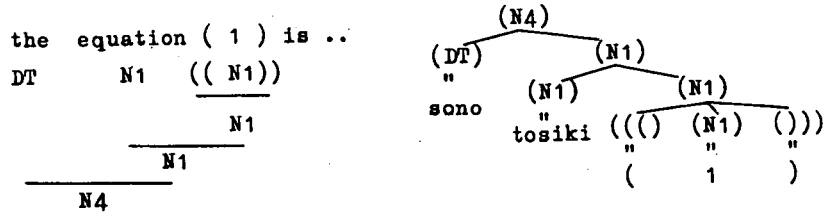
to table 1 and 2 in memory　　L1 · PA → L1 (2·1 , no·∅)
L1 PA

,,·Bl·,, ⟶ B2 (1·2·3 , ∅·∅·∅) : This rule is used in such a sentence as

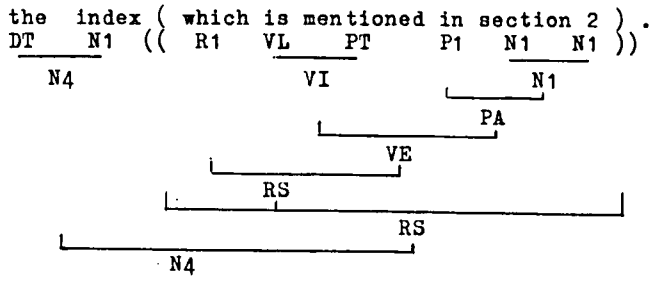


B2 corresponds to an adverbial phrase.

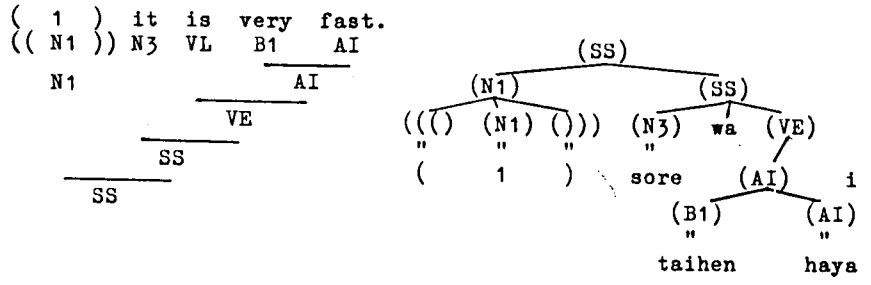
((· Nl ·)) ⟶ Nl (1·2·3 , ∅·∅·∅) : In scientific papers brackets appear frequently. This rule is applied as follows.



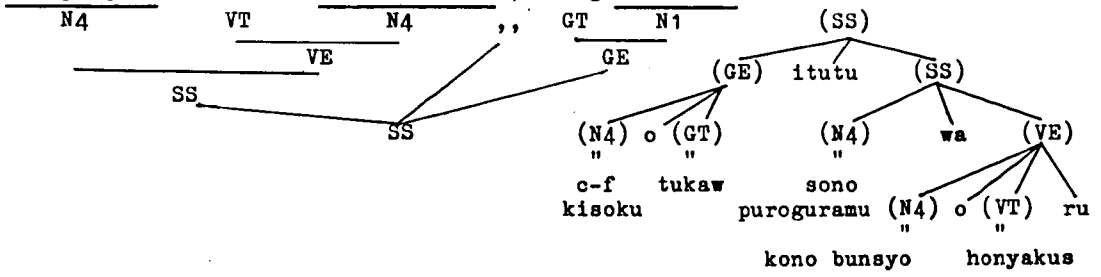
A more complex case can be processed even by the context-free rule.



The next form also occurs frequently.



SS,,GE ⟶ SS (3·2·1 , ITUTU·Ø·Ø) : This rule processes the case of participial construction.



There are many other patterns, which are listed in Table 4.5.1 in section 4.5.

One of defects may be that in this rewriting rule system it is

impossible to eliminate some sequence of symbol because the rewriting rule requires to be substituted always by one symbol. It is a question, however, whether such an operation is necessary or not, but when a given sentence is too complex to be treated by pattern, it may be one method to eliminate the ambiguous part in order to avoid mistranslation.

Some examples are shown, in Appendix-A (Fig. A-1 Fig. A-8), to show how sentences are parsed and synthesized.

Chapter 4

DICTIONARIES FOR MECHANICAL TRANSLATION

4.1 Word-ending processing

A word dictionary must be consulted one way or another in mechanical translation as well as in natural language processing. Though the conjugation of English and Japanese words is limited, it becomes a very important problem to study how to treat such conjugated words. That is, whether all paradigms are to be stored in the dictionary or not has a great influence on the processing method and efficiency. If word stems or original forms are to be stored as entries in the word dictionary by taking off the so-called inflection parts, the number of entries will greatly decrease. In this case, however, each conjugated word in the input sentence must be processed to the same stem form or original form as stored in the dictionary. But, generally speaking, it is difficult or may be even impossible to separate the real stem part from the inflection part by only one processing. For example, whether the word ending "ing" in the word "spring" is really an inflection or not is only determined after at least two trials. That is, if "spring" is a conjugated form of a verb, the stem or the original form is "spr" or "spre", so both possibilities must be searched in the dictionary. But in this case even if "spr" or "spre" is found in the dictionary, it can not be a desired word unless its part of speech is a verb. Because the inflection "ing" is possible only for a verb. On the other hand, if "spring" is not a conjugated word, the word "spring" itself must be searched. The reason why such tedious process is needed is that there are many cases in which one (spring) is not a

conjugated word but another (going) is a conjugated one, though both words have the same ending. The same holds in the cases of "-ed" and "-es". For example, the word "bed" is not a verb but it has a past-form ending, or the stem form of "machines" can not be determined by one trial because whether the word ending letter "es" is an inflection or can not be determined only from the morphological point of view. Therefore it may be easy to store stem forms or original forms, but it is difficult to find a stem form or an original form from a given conjugated word by mechanical procedure. Because, generally, the word dictionary for mechanical processing of a natural language contains a very large amount of information, however efficiently search algorithm may work, it takes much time to consult the dictionary several times per one word. Therefore it is necessary to design a more effective method which reduces the searching time.

The method which is described in this paper is simple but effective. Namely, if a word has one of such word endings as shown below, which

Table 4.1.1 Word-endings

E,	F,	S,	V,	Y,	
ED,	ER,	ES,	FE,	FS,	YS,
ERS,	EST,	FED,	FES,	IED,	IES,
ING,	SED,	SES,	VED,	VES,	YED,
VING					

may or may not be a real inflection suffix, the ending is separated from the word without any condition. If the remaining part has a duplicated consonant at the end of it and the taken-off ending are "-ed" or "-ing" or "er", the last letter (consonant) is cut off, for example, the word "running" first becomes "runn", and then the duplicated consonant is cut off and becomes "run". The remaining part which

is taken off the word ending, if any, is called the "pseudo-stem", and this pseudo-stem is registered in the dictionary as an entry. A few examples are shown. The pseudo-stem of "analyze" which has a word ending "e" becomes "analyz", and this form is registered. Now, if such words as "analyzes", "analyzed" or "analyzing" are given in the input text, by the end-processing the pseudo-stem of these words becomes "analyz"; therefore only one entry is necessary for four forms. As for the words "baby" and "babies", their pseudo-stem is "bab". The words "stop", "stops", "stopping" and "stopped" become "stop". In the case of "wife" and "wives", their pseudo-stem is "wi".

The unconditional separation of word endings which are not real inflection endings causes such cases where originally different words have the same pseudo-stem. Some examples are shown below.

A	B/E	CIT/E	SH/E
A/S	B/Y	CIT/Y	SH/ED
	B/ED		SH/Y
AN	B/EST	I	
AN/Y		I/F	TO
	BE/E	I/S	TO/E
APPL/E	BE/ER		TO/Y
APPL/Y	BE/ING	SHOULD	
		SHOULD/ER	

In these cases, they can not be distinguished only by the pseudo-stems; therefore each word ending which is taken off is considered when a word dictionary is consulted. The explanation of this procedure is described in section 5.1.

Sometimes there appear such words which have a word ending in their original form and further conjugate. For example, the word "enter" has the pseudo-stem "ent", and this form is registered in the dictionary.

Then if the input words are "enter" and "enters", their pseudo-stem becomes "ent", and so they can be found in the dictionary. But if the input words are "entered" and "entering", their pseudo-stem becomes "enter" and this does not coincide with the registered entry "ent". Therefore in this case both "ent" and "enter" must be sotred as entries for the verb "enter". Similar examples are shown.

DISCOVER ⟶ DISCOV + ER	INFER ⟶ INF + ER
DISCOVERING ⟶ DISCOVER + ING	INFERING ⟶ INFER + ING
INVEST ⟶ INV + EST	NEED ⟶ NE + ED
INVESTED ⟶ INVEST + ED	NEEDED ⟶ NEED + ED

This double registration of the same word can be avoided if word endings are allowed to be taken off repeatedly. For example, "entered" becomes "enter+ed" and then "ent+er+ed"; therefore its pseudo-stem becomes "ent". Then only one entry "ent" is sufficient for "enter", "enters", "entered" and "entering". By this method, however, the number of word pairs having some pseudo-stems increases, and to avoid the ambiguity, a large memory than a double registration is necessary. For example, "see", "sees", "seeing", "sing", "singing", "sings", "seed" and "singer" are constructed only by the word endings, so their pseudo-stem disappear by the repeated end-processing. If, however, only one ending is separated, the next four stems are needed.

S (SING, SINGS) , SE (SEE, SEES, SEEING) ,
SE (SEED) , SING (SINGING, SINGER)

Word-endings whose last letters are D, G, T, and R serve as morphemes which give some information as to parts of speech. That is, D suggests that if a word which has this word endings is a verb, it is a past or past participle form, and G suggests that it is an ing-form

verb, T and R suggest that in the case of an adjective it is comparative or superlative degree. But G and D do not give any information to other parts of speech except verbs, and R or T is only effective in the case of adjectives. Therefore the word ending "ed" in the word "bed" does not serve as a morpheme, because the part of speech of "bed" is not a verb.

As for the words which are not registered in the dictionary, their word endings are useful in infering parts of speech. That is, if their endings are D or G class, they are probably past form verbs or ing-form verbs, and if their ending is "ly", they may surely be adverbs.

Now, the following discussion concerns the conjugation of Japanese words. In Japanese verbs and adjectives conjugate, but it is too troublesome to give each word the information about the conjugation type. Then, a simple processing is applied to verbs and adjectives before they are registered in the dictionary. The processing is as below.

(1) As for the five-conjugation-type verbs (GODAN KATUYO), such as KAKU (KAKANAI, KAKIMASU, KAKU, KAKE BA, KAKO) or YOMU (YOMANAI, YOMIMASU, YOMU, YOMEBA, YOMO) etc., the word ending U is taken off from the third conjugation form. For example, KAKU and YOMU become KAK and YOM , and these stems are stored in the dictionary. But as a special case if the last syllable is constructed by only one vowel U , for example, KA/U or O/MO/U , the last sylable U is substituted with WU and then the word ending U is taken off. Therefore KAU and OMOU become KAW and OMOW . Another exceptional case is that such words whose last syllable is SU are stored in the dictionary without any processing. They are TOBASU (fly), KOROSU (kill), DAMASU (cheat) etc. It is to distinguish from the words in the case

of (4).

(2) As for the single-conjugation type words (KAMI ICHIDAN KATUYO, SHIMO ICHIDAN KATUYO), such as MIRU (MINAI, MIMASU, MIRU, MIREBA, MIYO) or SHIMERU (SHIMENAI, SHIMEMASU, SHIMERU, SHIMEREBA, SHIMEYO), the word ending RU is taken off from the third form, and stored in the dictionary. For example, MIRU and SHIMERU become MI and SHIME respectively.

(3) As for the irregular verb KURU (KONAI, KIMASU, KURU, KUREBA, KOI) which corresponds to English "come", the word ending RU is taken off, therefore KURU becomes KU .

(4) As for another irregular verb SURU (SINAI, SIMASU, SURU, SUREBA, SEYO) and its family words, the word ending URU is taken off. This type of word includes many compound words which have such a structure as "noun+SURU", for example, BENKYOSURU , KENKYUSURU (study), or SEISANSURU (produce) etc. They become BENKYOS , KENKYUS , or SEISANS .

(5) As for adjectives or adjectival words, the word endings I , NA , or NO is taken off from the modification forms. For example, UTUKUSII HANA (beautiful flower) becomes UTUKUSI , and SIZUKANA UMI (quiet sea) becomes SIZUKANA , and IKUTUKANO TAMAGO (some eggs) becomes IKUTUKA .

The above mentioned rules from (1) to (5) are used by man when he stores the translation words for verbs and adjectives in the dictionary. It is very easy for Japanese people to distinguish conjugation type of verbs and to separate stems from word endings. He can do it even by punching the words into computer.

On the other hand, the rules of conjugating these processed words, which will be explained in detail in Section 5.1, are very simple. In

short, if the last letter of a stem is a vowel, the inflection part is slightly shifted to begin with a consonant. On the contrary, if a stem ends in a consonant, the inflection part is make to begin with a vowel. For example, when the inflection RU is to be connected to ASOB (the stem form of ASOBU), the inflection part is changed to U because the stem ends in a consonant B , and the result is ASOBU .

If the stem is OSIE , and the inflection is RU , it becomes OSIERU because OSIE ends in a vowel, so the inflection must be begun with the consonant RU . When the past form inflection IMASITA is to be connected to YOM (stem of YOMU), it becomes YOM IMASITA because YOM ends in a consonant, and so the inflection part must be begun with a vowel. In the case of MI (stem of MIRU), it becomes MI MASITA . As for the special stems which have —SU or —KU forms, they are considered by the special routine.

The translation word which correspondsto the auxilially verb "will" or "shall" is stored in the form DAROW (stem of DAROU). Because if RU is connected to it, it becomes DAROWU and this Roman letters are translated into KANA letter as ダロウ . In this connection, it must be noted that the vowel after an assimilated sound is confusable with the syllables NA, NI, NU, NE, NO, so such single vowel (*) must be rewritten by W* form, and the syllable W* is transformed into KANA form in the same letter with a single vowel (*). For example TANI (unit) must be stored in the dictionary in a form TANWI , and when it is expressed in KANA letters, it becomes タンイ .

4.2 Word compression by cut-sum method

English words are constructed by the indefinite number of letters (usually from 1 to 16). If such words of varying length are stored in the dictionary as they are, it takes much time to search them and a large memory to store them. This becomes more serious as the number of words in the dictionary increases. Therefore some good methods which convert words of indefinite length to those of constant length must be studied. A calculation-address method is only effective when number of words is rather small and there are many core memories. The abbreviation method is not practical since its algorithm is complex. The effectiveness of tree like storage method is doubtful in reducing the memory size. Then the most simple and effective method which will be called the "cut-sum" method is to cut a word by every n letters and to add each segment simply looking like numerals. For example, the word "segmentation" is cut by every 5 letters, segme/ntati/on, and these segments are added, each segment looking like the binary number of 30 figures, as shown below.

S E G M E		110010 010101 010111 100100 010101
N T A T I		100101 110011 010001 110011 011001
+ O N	+	100110 100101 000000 000000 000000
		111110 101101 101001 010111 101110

The result is 5 character length ignoring overflowing figures. There are, however, some risks that originally different words may become the same head by this cut-sum method. The possibility of the risk differs according to the number of letters in each segment. For about eight thousand English words which the author used in the translation system, there appeared 274 pairs which become the same head by three letter cutting, and 26 pairs by four letter cutting. Some examples

are shown below.

LOOP POOL NOON ('OO)	IMPORTANT TRANSLATION (AU△)	CAT CONVERGENT (CAT)	TOUR ROUT/E ()OU)	NARRATION VARIATION (Q'1)
COURS/E SOURC/E (5OUR)	MATERIAL MISERABL/E (△¥4Y)	SHEEP WHEEL (IHEE)	HIGH NIGHT (HIGH)	COLLID/E GALLER/Y (4@LL)

By the five letter cutting, no same head appeared. Then in this English-Japanese translation system, the five letter cutting method is adopted. The possible number of words which can be constructed by five letters (30 bits) are $2^{30} \fallingdotseq 10^9$, and it is quite enough for English words.

The characteristic feature of this cut-sum method is that it is simple, and the same heads do not appear even by the five-letter cutting, and it needs a comparatively small number of bit per word to store. But this compression is non-reversible, that is, the cut-summed word can not be restored to the original form. But this is not a defect in the ordinary use of dictionary because the head word is only used to check if it coincides with the input word, and it is the information about parts of speech and translation words which are really wanted.

The compression is performed on the word whose word ending is processed by the method mentioned in Section 4.1.

4.3 Idiom dictionary

One of the obstacles to translation between natural languages

Table 4.3.1 Examples of idiom dictionary

1ST WORD	2ND WORD	MAIN WORD	4TH WORD	SUBSTITUTE	FUNCTION	CORRESPONDING-JAPANESE
DO	THE	SIGHTS	OF	AJRQ	VT	KENBUTUS
IS	NOT	TOO	MUCH	AJSQ	B1	OOKUNAI
AJSQ	TO	SAY	0	AJTQ	VL	KAGONDE WA ARAN
WILL	YOU	PLEASE	0	AJUQ	B1	DOOKA
IF	IT	WERE	NOT	AJVQ	B1	NAKEREBA
0	IN	SPITE	OF	AJYQ	P2	NIMOKAKAWARAZU
0	0	DISTINGUISH	WITH	AJZQ	VT	KUBETUS
0	0	REFER	TO	AKAQ	VI	GENKYUS
0	0	TOGETHER	WITH	AKBQ	P2	TO TOMONI
THAT	IS	TO	SAY	AKCQ	B1	SUNAWACHI
HOW	DO	YOU	DO	AKDQ	SS	IKAGADESU
0	IS	FOND	OF	AKEQ	VT	SUK
0	IN	TURN	0	AKFQ	B1	JUNNI
0	AND	SO	ON	AKGQ	B1	NADO
0	0	SEEK	FOR	AKHQ	VT	SAGAS
0	0	APEEK	OF	AKIQ	VT	KATAR
HAVE	NO	USE	FOR	AKJQ	VT	HITUYO TO SIN
0	0	NEED	NOT	AKKQ	VA	* NIOYOBAN
0	AS	SOON	AS	AKLQ	WN	YA INAYA
0	0	THANK	YOU	AKMQ	B1	ARIGATO
0	AS	LONG	AS	AKNQ	WN	KAGIRI
0	SO	LONG	AS	AKOQ	WN	NARABA
MOST	OF	US	0	AKPQ	N4	WAREWARE NO OOKU
0	0	BORNE	OUT	AKQQ	VA	*U BEKIDE AR
0	0	OUGHT	TO	AKQQ	VA	*U BEKIDE AR

Table 4.3.1 Examples of idiom dictionary

1ST WORD	2ND WORD	MAIN WORD	4TH WORD	SUBSTITUTE	FUNCTION	CORRESPONDING-JAPANESE
0	IN	VIEW	OF	INV	P2	NO TENKARA
AS	IT	WERE	0	IWABA	B1	IWABA
0	0	ASMAT	FACT	JISAI	B1	JISSAI
0	0	KNOW	OF	KNOW	VT	SIR
0	0	LOOK	OUT	LOOK	VT	MI
0	0	MADE	UP	MADE	PD	TUKUR
0	0	MAKE	UP	MAKE	VD	TUKUR
0	0	MAKE	UP	MAKE	VD	TUKUR
0	0	PLENTY	OF	MANY	AO	OOKU
0	AN	ABUNDANCE	OF	MANY	AO	OOKU
0	0	MANY	A	MANY	AO	OOKU
0	0	DOZENS	OF	MANY	AO	OOKU
0	0	LOT	OF	MANY	AO	OOKU
MAY	AS	WELL	0	MAY	VA	*U KAMOSIREN
0	CANNOT	CHOOSE	BUT	MUST	VA	*E NEBANARA
0	0	MUST	NEEDS	MUST	VA	*E NEBANARA
0	0	HAVE	TO	MUST	VA	*E NEBANARA
0	0	HAD	TO	MUST	VA	*E NEBANARA
0	HAVE	GOT	TO	MUST	VA	*E NEBANARA
OUGHT	TO	HAVE	0	MUST	VA	*E NEBANARA
0	NOT	ONLY	0	NONL	B1	TANNI
0	0	'	T	NOT	EE	%N
0	NEW	YORK	0	NYK	N1	NYUYOOKU
0	OF	COURSE	0	OBUKO	B1	MOTIRON
FELL	IN	LOVE	WITH	RENAI	PI	HORE

Table 4.3.1 Examples of idiom dictionary

1ST WORD	2ND WORD	MAIN WORD	4TH WORD	SUBSTITUTE	FUNCTION	CORRESPONDING-JAPANESE
0	BY	TURNS	0	AHMQ	B1	KAWARUGAWARU
HAVE	NO	USE	FOR	AHNQ	VT	HITUYOTO SIN
AFTER	A	WHILE	0	AHOQ	B1	SIBARAKUSITE
AS	A	WHOLE	0	AHPQ	B1	ZENTAITOSITE
WORD	FOR	WORD	0	AHQQ	B1	1GO1GO
0	A	DAY	0	AHRQ	B1	1NICHINI
BEST	OF	ALL	0	AHSQ	B1	ICHIBAN
0	WAS	BORN	0	AHTQ	P1	UMARE
0	0	CAME	OUT	AHUQ	P1	DETEK
THANK	YOU	VERY	MUCH	AHVQ	B1	ARIGATO
THE	DAY	BEFORE	YESTERDAY	AHWQ	B1	OTOTOI
0	FAR	AWAY	0	AHXQ	A1	TOO
0	0	WAKE	UP	AHYQ	P1	OKI
0	0	LOOK	ABOUT	AIAQ	V1	SAGAS
0	0	CONSULT	WITH	AIBQ	V1	SODANS
0	0	COME	BACK	AICQ	V1	MODO
0	0	STAND	UP	AIDQ	V1	TAT
HERE	AND	THERE	0	AIEQ	B1	KOKOKASIKO
UP	AND	DOWN	0	AIFQ	P1	NO ACHIKOCHI
0	AT	ALL	0	AIGQ	B1	SUKOSIMO
AS	YOU	KNOW	0	AIHQ	B1	GOZONJINOYONI
0	YEAR	AFTER	YEAR	AIIQ	B1	NENNEN
0	0	PUT	UP	AIJQ	V1	TOMAR
OVER	AND	OVER	AGAIN	AILQ	B1	NANKAIMO
0	WOULD	LIKE	TO	AIMQ	VA	*U KOTOO HOS

Table 4.3.1 Examples of idiom dictionary

1ST WORD	2ND WORD	MAIN WORD	4TH WORD	SUBSTITUTE	FUNCTION	CORRESPONDING-JAPANESE
0	IN	CONCLUSION	0	AFNQ	B1	SAIGONI
0	IN	CONTACT	WITH	AFOQ	P2	TO SESSYOKUSITE
ON	THE	CONTRARY	0	AFPQ	B1	HANTAINI
0	OF	COUSE	0	AFQQ	B1	MOTIRON
0	OF	COUSE	0	AFQQ	B1	MOTIRON
A	MATTER	AFQQ	0	AFRQ	N1	TOZENNOKOTO
AS	A	MATTER	AFQQ	AFSQ	B1	TOZEN
OUT	OF	DATE	0	AFTQ	A0	KYUSIKI
THIS	DAY	WEEK	0	AFUQ	B1	HAISHUNO KYU
ON	THE	DECREASE	0	AFVQ	B1	GENSYOSITE
0	WITH	EASE	0	AFWQ	B1	YUINI
0	FOR	INSTANCE	0	AFXQ	B1	TATOEBA
0	FOR	EXAMPLE	0	AFXQ	B1	TATOEBA
AT	THE	EXPENSE	OF	AFYQ	P2	O DASINISITE
TO	SOME	EXTENT	0	AFZQ	P2	ARUTEIDO
0	IN	FACT	0	AGAQ	B1	ZISSAI
0	AT	FIRST	0	AGBQ	B1	HAJIMEWA
0	ON	FOOT	0	AGCQ	B1	ARUITE
ON	THE	GROUND	OF	AGEQ	P2	NO RYUDE
0	BY	HEART	0	AGFQ	B1	SOREDE
0	IN	HONOR	OF	AGGQ	P2	NO KINENNI
IN	A	HURRY	0	AGHQ	B1	AWATETE
MAKE	A	JOURNEY	0	AGIQ	V1	RYOKO O S
MADE	A	JOURNEY	0	AGJQ	P1	RYOKO O S
0	AT	LAST	0	AGKQ	B1	TUINI

Table 4.3.1 Examples of idiom dictionary

1ST WORD	2ND WORD	MAIN WORD	4TH WORD	SUBSTITUTE	FUNCTION	CORRESPONDING-JAPANESE	
HAPPY	NEW	YEAR	0	AQAQ	SS	SINNEN OMEDETOO	
0	0	'	R	ARE	VL	AR	
0	0	ARRIVE	AT	ARRIVE	VI	TOTYAKUS	
0	AT	PRESENT	0	ATP	B1	MOKKA	
0	0	BECAUSE	OF	BEKO	P2	NO TAME	
0	BY	DINT	OF	BY	P2	NI YOTTE	
0	BY	MEAN	OF	BY	P2	NI YOTTE	
0	BY	MEANS	OF	BYM	P2	NI YOTTE	
0	CALL	FOR	0	CALL	VI N1	YOB	YOBIGOE
0	BE	ABLE	TO	CAN	VA	*U KOTOGADEKI	
0	0	COME	OVER	COME	VI	KU	
CONGRATURATION	ON	YOUR	0	CONGRA	P2	OMEDETOO	
0	0	CUT	OFF	CUT	VI AI	KIR	KIRA
0	0	DEALT	WITH	DEALT	PT	ATUKAW	
0	0	EQUIVALENT	TO	EQVIO	P2	NI HITOSII	
0	0	EVERY	DAY	EVRDA	B1	MAINITI	
0	A	BIT	OF	FEW	AO	SUKOSI	
0	0	OUT	OF	FROM	P2	KARA	
AS	A	GENERAL	RULE	GENERALLY	B1	TAITEI	
0	IN	GENERAL	0	GENERALLY	B1	TAITEI	
0	GO	OUT	0	GO	VI N1	IK	IKU
YES	,	I	DO	HAIIQ	SS	HAI	
0	MAKE	HASTE	0	HURRY	VI	ISOG	
0	IN	SHORT	0	INSY	B2	SUNAWATI	
,	THAT	IS	,	INSY	B2	SUNAWATI	

which belong to different language families is that there are many characteristic expressions in each language, in which it is very difficult to grasp the whole meaning of expression from their constituents. It is, as it were, exocentric structure in meaning. Such expressions do play an essential role in natural languages, so it is a very important problem to study how to translate them. But there is no general algorithm to translate idiomatic expressions as well as ordinary structures by means of the same procedure. There is no choice but to translate the whole idiomatic expression into another phrase using one-to-one correspondence table. Of course there is not, in the strict sense, any conectly one-to-one correspondence, but partially equivalent expressions can be found out. The idiom dictionary contains such one-to-one correspondent expressions.

Idiomatic phrases usually consist of several words, say, from two to four, and they very rarely consist of more than six words. They are supposed here to be constructed by continued sequence of words, and discontinuous expression (correlative one) such as "not...but" or "so...that" etc. are not treated because they are difficult to substitute with one word and to process with fixed pattern matching.

It is a question how idioms are to be stored in the dictionary. The number of words which constitute idioms is not definite, then to make it easy to search idioms it is desirable to convert them into regular forms having the same length. But unlike the word dictionary such a simple compression method is not effective with the idiom dictionary, because the elements of idioms are words carrying meanings and a simple cut-sum method loses much information involved in each word. Therefore the following method is adopted to make it possible to understand its forms easily and also to make the program simple and

quick, though it takes rather redundant memories.

Usually an idiomatic expression contains a word which characterizes a phrase. Take for example, such words as "spite", "advantage" or "rate" in the next expression "in spite of", "take advantage of" or "at any rate". Such words generally appear less frequently; therefore it is efficient to set them in such a position that they are first looked for. Most idioms have the structures like CCM, CMC, CM, MC, and CCMC, where M is a main word and C is the other word in an idiomatic expression. Then they are arranged so that the main words are situated in the third position, as shown below,

1st	2nd	main	4th	substitution
C	C	M	C	S
-	C	M	C	S
-	-	M	C	S
-	C	M	-	S
C	C	M	-	S

Fig. 4.3.1 Structure of idiom dictionary.

Except CCMC they contain redundant memories (slashed part), but by putting them, the search program becomes easy to make. As for the idioms which consist of more than five words or have such structure as CMCC or MCC, they are stored by dividing them into several parts. For example, the set phrases "as a matter of fact" or "not to much to say" are stored in the next forms.

1st	2nd	main	4th	substitution
A/S	A	MATT/ER	O/F	QXQ
-	-	QXQ	FACT	QXYQ
I/S	NOT	TOO	MUCH	QZQ
QZQ	TO	SA/Y	-	QZYQ

A word which substitutes the idiom is situated in the fifth position. This word is not necessarily a usual English word, but it is, in most cases, a sign to distinguish idioms from other words, and registered in the word dictionary. The word QXYQ in the above example is registered in the word dictionary together with other usual words as below (Fig. 4.3.3).

original form	head	part of speech	Japanese
BOY	BOOOO	N1	SYONEN
GONE	GONOO	PI	IK
QXYQ	QXYQO	B1	JISSAI
QZYQ	QZYQO	VI	IISUGIDE ARAN

Fig. 4.3.3 word dic. for idiom equivalence.

Each word in the idiom is in its pseudo-stem form and compressed into five characters by the cut-sum method mentioned in Section 4.2. Therefore one idiom is stored using 5 X 5 characters in the memory. Several examples are shown in Table 4.3.1 together with all information which is contained in the word dictionary.

4.4 Word dictionary

The word dictionary is the most important dictionary in the mechanical processing of natural languages, by which several kinds of information is given to each input word according to its purpose of processing. The information which must be included in the word dictionary for mechanical syntactic translation from English into Japanese is English head words, corresponding Japanese, and grammatical informa-

tion about English words. Grammatical information about Japanese is not necessary in the syntax-to-syntax translation treated in this paper because almost all information is taken into consideration in the syntactic rules and the conjugation of verbs and adjectives in Japanese is recognized by investigating the last letter of verbs and adjectives. No semantic information is given in the present system.

The memory necessary for one word is 13 characters plus the number of letters in Japanese words. The stored form is shown below with its original form included in the ordinary English-Japanese dictionary.

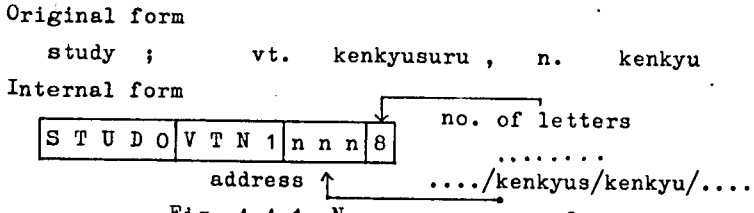


Fig. 4.4.1 Necessary memory for a word in the word-dictionary.

Five characters are used for the compressed word head, and four characters are reserved for the two parts of speech, two characters for each. The three characters indicate the top address of Japanese words. The second translation words, if any, are indicated by the relative address to the first one by one character. The reason why translation words are stored in a different place and linked by a pointer is simply to express each entry in a regular form and to reduce the number of steps to consult the word dictionary.

All entries, about 8000 words, are sorted in ascending order taking the pseudo-stems of five characters as key words. There are, however, several words which have the same pseudo-stems on account of word-end processing. In this case correct information can not be obtained. Therefore such words are registered in a special manner.

That is, the word ending which is taken off is stored instead of the pseudo-stem except for the first head among the same head words. For example, these words "I", "IF", and "IS" all become "IØØØØ" by the end processing and cut-sum compression. If they are arranged simply in ascending order, the same head words appear in succession. Then the

(original form)	head	:	(original form)	head
I	I0000		TO	T0000
IF	0000F		TOE	0000E
IS	0000S		TOY	0000Y
			TOYS	000YS

Fig. 4.4.2 Arrangement for the same head words

second and third head words are substituted by their endings as shown in Fig. 4.4.2. The endings are stored from the least significant character and arranged in ascending order. In this case, however, it is a question which of them must be put in the first position. Though this depends on the search algorithm, generally it is reasonable to put such word as conjugates. Because the first head word among the same head words is not given its word ending, conjugated words coincide with the head if their endings are taken off, if verbs are not stored as shown in Fig.4.4.3. But if verbs of frequent occurence are stored in the first position, the number of entries can be reduced.

head		:	head	
LI000	(lie)		LI000	(live,lived,lives,living)
000FE	(life)		0000E	(lie)
000VE	(live)		000FE	(life)
00ING	(living)			
00VED	(lived)			
00VES	(lives)			

Fig. 4.4.3 Comparison between two arrangements

Table 4.4.1 Samples of word dictionary

ENGLISH	SYMBOL	JAPANESE	HEAD	ENGLISH	SYMBOL	JAPANESE	HEAD
ASHAMED	N1	HAZITE	ASHAM	AVERSE	AO	KIRATTE	AVERS
ASHORE	B1	RIKUZYONI	ASHOR	AVOID	VT	SAKE	AVOID
ASIA	AO	AZIA	ASIAO	AWAY	B1	HANARETE	AWAOO
ASIAN	AO N1	ASIA/ASIAJIN	ASIAN	AWAITE	VT	MAT	AWAIT
ASIDE	B1	WAKINI	ASIDO	AWAKE	VI AI	MEZAME/NEMURANA	AWAKO
ASK	VT	TAZUNE	ASKOO	AWARE	N1	SITTE	AWARO
ASPIRE	VI	NETUBOS	ASPIR	AWARD	N1 VT	SHO/ATAE	AWARD
ASSUME	VT	HURIOS	ASSUM	AWFUL	AI	OSOROSI	AWFUL
ASSURE	VT	TASIKAME	ASSUR	AWHILE	B1	SIBARAKU	AWHIL
ASTRAY	B1	MITINIMAYOTTE	ASTRA	AX	N1 VT	ONO/NATAOHURU	AXOOO
ASUNDER	B1	HANARETE	ASUND	AXIS	N1	JIKU	AXIOO
AT	P1	NO TOKORO	ATOOO	AXLE	N1	SINBO	AXLOO
ATE	PT	TABE	OOOOE	ASSIMILATION	N1	DOKA	A,4:%
ATOM	N1	GENSI	ATOMO	BE	VL	AR	BOOOO
ATP	B1	MOKKA	ATPOO	BY	P2	NI YOITE	OOOOY
COMMEMORATION	N1	KINEN	ATSW8	BED	N1	NEDOKO	OOOED
ATTER	VI N1	PARAPARAFUR/PARAPARAO	ATTOO	BEST	RA	MOTTOMOYO	OOEST
ATTEST	VT	SHOMEIS	OOEST	COMBINATORIAL	AO	KUMIAWASE	B9@Z2
ATTIRE	VT N1	SEISOSAS/YOSOOI	ATTIR	RELINQUISH	VT	YAME	B9%=%
AUNT	N1	OBA	AUNTO	INFLUENTIAL	AN	EIKYOIEKI	B=9%5
JUSTIFICATION	N1	SEITOKA	AUY5:	INSIDENTALLY	B1	FUZUITEKINI	B=N#X
CONTEMPLATION	N1	TINSIMOKKO	AU#58	BAY	N1 VT	WAN/HOE	BAOOO
AUTO	N1	JIDOSHA	AUTOO	BASES	N1	KISO	OOSES
AVAIL	VI N1	YAKUNITAT/RIEKI	AVAIL	BABY	N1	AKANBO	BABOO
AVENGE	VT	KATAKIOUT	AVENG	BABBLE	VI N1	OTO O TATE/SESERAGI	BABBL

Table 4.4.1 Samples of word dictionary

ENGLISH	SYMBOL	JAPANESE	HEAD	ENGLISH	SYMBOL	JAPANESE	HEAD
CORE	N1	SIN	COR00	CRAM	VT N1	TUMEKOM/SUSIZUME	CRAM0
CORAL	N1 AO	SANGO/SANGO	CORAL	CRANE	N1 VI	TURU/KUBIONOBAS	CRAN0
CORD	N1	TUNA	CORD0	CREED	N1	SINJO	CREE0
CORN	N1	KOKUMOTU	CORN0	CREAM	N1	KURIMU	CREAM
CORNER	N1	KADO	000ER	CREATE	VT	SOZOS	CREAT
CORPS	N1	GUNDAN	CORP0	CREEK	N1	IRIE	CREEK
CORPSE	N1	SITAI	CORPS	CREEP	VI N1	HA/HAUKOTO	CREEP
COSMOS	N1	UCHU	COSM0	CRIF	N1	DANGAI	CRI00
COST	VT N1	YOS/HIYO	COST0	CRIB	N1	SHOINYOSHINDAI	CRIB0
COSTLY	AN	KOKA	COSTL	CRIME	N1	TUMI	CRIM0
COSUL	N1	RYOZI	COSUL	CRISIS	N1	KIKI	CRISI
COULD	PU	*U KOTOGADEKI	COULD	CROOK	N1	MAGATTAMONO	CROOK
COUNTER	N1	KAZOERUMONO	COUNT	CROSS	VT N1	YOKOGIR/JYUJI	CROS0
COUPLE	N1	ITTUI/TUNAG	COUPL	CROW	N1 VI	KARAS/NAK	CROW0
COURSE	N1	KOOSU	COURS	CROWED	N1 VI	GUNSYU/MURAGAR	000ED
COURT	N1 VT	KYUTEI/AIOMOTOME	COURT	CROWD	N1 VI	GUNSYU/OSISUSUM	CROWD
COVER	VT	OOW	COVER	CROWN	N1 VT	OKAN/OINITUKASE	CROWN
COVET	VT	HOSIGAR	COVET	CRUDE	AN	SOMATU	CRUD0
COW	N1 VT	MEUSHI/ODOKAD	COW00	CRUISE	VI N1	JUNKOS/JUNKO	CRUIS
COZY	AI	IGOKOTINOYO	COZ00	CRUISER	N1	JUNYOKAN	000ER
CRY	VT N1	SAKEB	CR000	CRUMB	N1	PANKUZU	CRUMB
CREST	N1	TOSAKA	00EST	CRUSH	VT N1	OSITUBUS	CRUSH
CRAVE	VI	KONGANS	CRA00	CRUST	N1 VT	PANNOKAWA/GAIHIDEOOW	CRUST
CRAB	N1	KANI	CRAB0	CRVEL	AN	ZANNIN	CRVEL
CRAFTY	AN	KOKATU	CRAFT	MANIFESTATION	N1	MEIJI	C*&Y9

Table 4.4.1 Samples of word dictionary

ENGLISH	SYMBOL	JAPANESE	HEAD	ENGLISH	SYMBOL	JAPANESE	HEAD
SEMICOLON	N1	SEMIKORON	HZ' D	OUTLET	N1	DEGUTI	IUTLE
SECTION	N1 VT	SETUDAN/KUBUNS	H@CTI	OUTPUT	N1	SYUTURYOKU	IUTPU
MANCHURIA	N1	MANSYU	H@ ED	IVY	N1 VT	TUTA/TUTADEOO	IVOOO
TOUGHNESS	N1	YOJINSA	H(OGH	IVORY	AO	ZOGE	IVORO
VISUALLY	B1	SIKAKUTEKINI	H(SUA	IWABA	B1	IWABA	IWABA
SCISSORR	N1	HASAMI	H%2SS	NEEDFUL	AN	HITUYO	IYEDF
MIRACULOUS	AN	KISEKITEKI	H%+5C	TENSION	N1	KINCHO	I@NSI
MINIMUM	N1	SAISYO	H%NIM	PROPOSAL	N1	MOSIKOMI	I,9PO
SHAMPOO	N1 VT	SENPATU/KAMIOARAW	H ₹₹₹	TEASPOON	N1	CHASAJI	I,WSP
SEASHORE	N1	KAIGAN	H ₹₹?	TEXTBOOK	N1	KYOKASYO	I(I1B
SPENDTHRIFT	N1 AI	ROHIKA/KANEZUKAINOARA	H$	CHARACTERISTIC	N1	TOKUCYO	I$0N#
I	N4	WATASHI	I0000	CHARACTERISTICS	N1	YOKUCYO	0000S
IF	C2	NARA	0000F	PHILOSOPIC	AO	TETUGAKU	I$0(Z
IS	VL	AR	0000S	POSSESIVE	AO	SYOYU	I$55E
OPPRESIVE	AN	ASSEITEKI	I0PRE	OPERATION	N1	SOOSA	;0(>A
MANOEUVER	N1 VT	ENSYU/ENSYUS	I6NOE	PROVISION	N1 VT	JUNBI/KYOKYUS	;3 ;I
TERMINUS	N1	MOKUTEKITI	I9RMI	MATHEMATICAL	AO	UGAKU	;60/0
UNCOMMON	AN B1	HIBON/MARENI	I=YOM	PERMISSION	N1	KYOKA	;83*
UNANIMOUS	AO	MANZYOITTI	I:5NI	PERSIST	VI	KOSITUS	;8RSI
UNANIMOUSLY	B1	IKUCHIDOONNI	I:6G(	PERSISTENCE	N1	KOSITU	;8$G#
SHALLOW	AI N1	ASA/ASASE	I>ALL	PERSISTENCY	N1	KOSITU	0000Y
NUTRITION	N1	EIYO	I>$>I	PENINSULAR	AO	HANTO	;'8Y>
SPONSOR	N1 VT	SPONSOR/KOENS	I+ONS	PURPOSELY	B1	WAZARO	;':PO
VACUUM	N1	INKU	IACUU	SMALLPOX	N1	TENNENTO	;=8LL
MANEUVER	N1 VT	ENSYU/ENSYUS	IANEU	PIANIST	N1	PIANISUTO	;:ANI

Table 4.4.1 Samples of word dictionary

ENGLISH	SYMBOL	JAPANESE	HEAD	ENGLISH	SYMBOL	JAPANESE	HEAD
INFLUENZA	N1	INHURUENZA	I=&UU	FAMILIARITY	N1	SINKO	IL IF
INNOCENT	AN	MUJYAKI	I=HOC	FANATIC	AN N1	NEKKYOTEKI/KYOSINSHA	IMNAT
ENVIRONMENT	N1	KANKYO	I=II>	ETERNITY	N1	EIEN	IOERN
INSOLENT	AN	BUREI	I=NOL	SURPRISINGLY	B1	ODOROKUHODONI	IP3 0
EXPEDIENCE	N1	BENGI	I :QD	TELEGRAPHIC	AO	DENSIN	IP'"<
EXPEDIENCY	N1	BENGI	0000Y	FRONTIER	N1	ZENSEN	IRONT
EXPERIENCE	N1 VT	KEIKEN/KEIKE	I :OR	GROWTH	N1	SEITYU	IROAT
EXPEDIENT	AN N1	KOTUGO/BENGISHUDAN	I 8D	EXHIBIT	VT N1	MISE/TENZI	I#HIB
IMMODERATE	AO	KADO	I WID	EXPLOIT	VT N1	KAIHATUS/TEGARA	I#PLO
IMMATERIAL	AN	HIBBUSITUTEKI	I SLF	EXPEDITION	N1	TANKEN	IY0•Z
INNUMERABLE	AO	MUSU	I>X77	EXPOSITION	N1	SETUMEI	IY1 G
INNUMERABLY	B1	MUGENNI	0000Y	FINACIAL	AO	ZAISEI	IY8AC
INTOLERABLE	AI	GAMANDEKINA	I64Z6	PERSEVERANCE	N1	NINTAI	IYC3e
INTORERAMCE	N1	GAMANDEKINUKOTO	I65'(	REVOLUTIONARY	AO	KAKUMEI	I56 8
EQUATION	N1	HOOTEISIKI	I6IAT	FUGITIVE	N1 AI	TOBOSHA/NIGETA	IUGIT
ERUPTION	N1	BAKUHATU	I+IPT	IMPCRFECT	AN	HUKANZEN	IZ•6R
ORTHONORMAL	AO	CYOKKO	I+)(X	IMPERFECTLY	B1	FUKANZENNI	IZ•9:
INTERESTED	PT	MIRYOS	IHOER	FEMININE	AO	ZYOSEI	I&MIN
INTERESTING	AI	OMOSIRO	00ING	INTERFERE	VT	KANSYOS	I•)ER
IMMEASURABLE	AO	HIKASOKU	II OL	FEROCIOUS	AN	DOMO	I(%UC
EXCLAIM	VT VI	SAKEB/KANTANS	I•CLA	FOOTLIGHTS	N1	KYAKKO	I5$UL
EXPERIMENT	N1	JIKKEN	I•(•)	FORESIGHT	N1	SENKENNOMEI	I 7JF
EXPLAIN	VT	SETUMEIS	I)PLA	FORTNIGHT	N1	NISYUKAN	I 7JE
EXTERIOR	AO N1	GAIBU/GAIBU	I)ER	INTENTIONAL	AN	KEIKAKUTEKI	I$IeN
FAMILIAR	AI	SITASI	IL IL	CONSERVATISM	N1	HOSHUSHUGI	I5XN#

Table 4.4.1 Samples of word dictionary

ENGLISH	SYMBOL	JAPANESE	HEAD	ENGLISH	SYMBOL	JAPANESE	HEAD
USE	VT	TUKAW	U5000	LOGICAL	AN	RONRITEKI	V9GIC
USA	N4	GASSYUKOKU	USA00	LITERATE	AI N1	KYOIKUNOARU/KYOYOSHA	V:TER
USABLE	AN	SIYOKANO	USABL	LIMITATION	N1	SEIGEN	V: ??
USAGE	N1	YOHO	USAG0	LITERATURE	N1	BUNGAKU	V P X
USHER	VT	MONEI/TORITUG	USH00	VAGUE	AN	AIMAI	VAGU0
ASSAIL	VT	SHUGEKIS	USSAI	VAIN	AN	MUEKI	VAIN0
USSR	N1	SOBIET	USSR0	VALVE	N1	EN	VAL00
USUAL	AO	FUTU	USUAL	VALID	AN	TASIKA	VALID
INAUGURATION	N1	SHUNINSIKI	UTLP<	VALLEY	N1	TANI	VALLE
UTPIA	N1	RISOKYO	UTPIA	VALUE	N1 VT	KATI/HYOKAS	VALU0
UTTER	AO VT	MAITAKU/KUTINIDAS	UTT00	VANIY	N1	KUKYU	VANI0
AUDIBLE	AO	KATYO	UUDIB	VANITY	N1	KYOEI	VANIT
JUSTICE	N1	SEIGI	UUSTI	VARY	VT	KAE	VAR00
AVOWAL	N1	KOGEN	UVOWA	BARREL	N1	TARU	VARRE
ACTUALLY	B1	HONTONI	UWTUA	VAST	AN	KODAI	VAST0
BLANCKET	N1 VT	MOHU/MOHUDETUTUM	UZ4NC	ACADEMY	N1	GAKKO	VCADE
LANDMARK	N1	KYOKAIHYOSIKI	U.7DM	BEGUILE	VT	DAMAS	VEGUI
LIBERAL	AN	ZIYU	U(BER	VEIL	N1	BERU	VEIL0
LIBERALISM	N1	JIYUSHUGI	U(*8	VEIN	N1	KEKKAN	VEIN0
LIBERALLY	B1	KANDAINI	U(VER	VERY	B1	TAIHEN	VER00
V5	P2	NI TAISITE	V0000	VERB	N1	DOSI	VERB0
VEST	N1	CHOKKI	00EST	VERDUE	N1	MIDORI	VERDU
LITERARY	AO	BUNGAKU	V2TER	VERGE	N1 VI	FUTI/MUKAW	VERG0
LABORATORY	N1	ZIKKENSITU	V4Z6R	VERIFY	VT	RISSYUS	VERIF
ANNUALLY	B1	NENNEN	V8NUA	VERSE	N1	INBUN	VERS0

Several examples of compressed words and their arrangement are shown in Table 4.4.1.

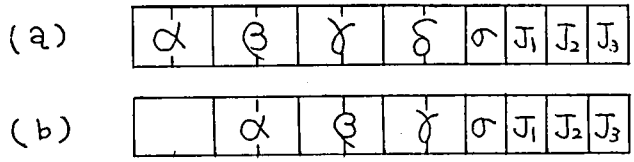
4.5 Syntactic dictionary

In section 3.5 several structures of syntactic rules are shown. Their internal forms are described here. The rewriting rules are divided into two classes according to the number of characters which are needed to store. They are 3-symbol rules (a) and 2-symbol rules(b).

(a) $\alpha \cdot \beta \cdot \gamma \longrightarrow \delta(\sigma, \xi)$

(b) $\alpha \cdot \beta \longrightarrow \gamma(\sigma, \xi)$

Their internal character construction is shown below.



The 3-symbol rule has 12 characters, that is, 6 characters for the English pattern, 2 for the substitution symbol, 1's for the word-order indicator and 1 for particles to be inserted. The 2-symbol rule has 10 characters: 4 characters for the English pattern, the remaining part is the same as the 3-symbol rule. There are 21 kinds of word order indicators (σ), and 63 different particles to be inserted.

The rewriting rules are classified into two major classes according to their syntactic hierarchy. This corresponds to (I) and (II) in

section 3.3. Each class is divided into sub-classes, that is, 3-symbol rules and 2-symbol rules. Therefore there are four classes, that is, (A-3), (A-2), (B-3) and (B-2) classes, and their hierarchy is in that order. Though there are only four classes, they are equivalent to at least eight classes because of the repeated application of them by the matching algorithm which will be explained in section 5.1.

All rewriting rules are listed in Table 4.5.1, but for the sake of easy understanding some editions are applied to the stored forms, that

Stored form

α	β	γ	δ	σ	J_1	J_2	J_3

σ = 2 1 3

(J_1)= 亻 , (J_2)= 口 , (J_3)= 丿乀

Listed form

PATTERN (1)	PATTERN (2)	PATTERN (3)	SUBSTITUTE	CORRESPONDING-JAPANESE
α	β	γ	δ	(2)/亻/ (1)/口/ (3)/丿乀/

is, instead of symbol expression, corresponding Japanese are explicitly expressed.

There are no rules which have the same pattern part ($\underline{\alpha\cdot\beta\,\gamma} \longrightarrow \delta$), and in each class the rules are sorted simply in ascending order with pattern part, because there is no difference in hierarchy between the rules in the same class.

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ••• A -- 3

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
=	B1	••	B2	(2)(3)
=	GE	••	B2	(2)/RU TO/(3)
=	L1	SS	SS	(2)/ E/(3)
=	N1	=	SS	(2)
=	N4	=	SS	(2)
=	PA	••	B2	(2)/ NI/(3)
=	PI	N1	N4	(2)/RARETA/(3)
=	SS	=	SS	(2)
=	WS	••	B2	(2)(3)
AI	AS	N4	AI	(3)/ TO ONAJI/(1)
AI	C1	AI	AI	(1)/KU/(2)(3)
AI	--	N1	N1	(1)/II/(2)(3)
AI	--	PT	N1	(1)/II/(2)(3)/IMASITA/
AI	PI	SS	SS	(1)/II/(2)/RARETA/(3)
AI	PT	SS	SS	(1)/II/(2)/RARETA/(3)
AN	AS	N1	AN	(3)/ TO ONAJI/(1)
AN	C1	AI	AI	(1)/ DE/(3)
AN	C1	N1	N1	(1)/NA/(2)(3)
AN	C4	N4	N4	(1)(2)(3)
AN	--	PI	N1	(1)/NA/(2)(3)/IMASITA/
AN	PI	N1	N1	(1)/NA/(2)/IMASITA/(3)
AN	••	AO	AO	(1)/NA/(2)(3)
AO	AS	N4	AO	(3)/ TO ONAJI/(1)
AO	C1	N1	N1	(1)/NO/(2)(3)
AO	--	N1	N1	(1)/NO/(2)(3)
AO	PI	N1	N1	(1)/NO/(2)/IMASITA/(3)
AO	••	SS	SS	(1)(2)(3)
B1	AN	N1	N1	(1)(2)/NA/(3)
B1	C1	B1	B1	(1)(2)(3)
B1	RA	N1	N1	(1)(2)/II/(3)
C1	B1	SS	SS	(1)(2)(3)
C3	N4	))	))	(1)(2)(3)
DT	B1	N1	N4	(1)(2)(3)
DT	B1	PT	DT	(1)(2)(3)/RARETA/
DT	GI	N1	N4	(1)(2)/ITUTUARU/(3)
DT	GT	N1	N4	(1)(2)/ITUTUARU/(3)
DT	PI	N1	N4	(1)(2)/RARETA/(3)
DT	RA	N1	N4	(1)/ YORI/(2)/II/(3)
EE	VH	PD	VD	(3)/IOE//AN/
EE	VH	PI	VI	(3)/IOE//AN/
EE	VL	GD	VD	(3)/ITUTU//AN/
EE	VL	GI	VI	(3)/ITUTU//AN/
EE	VL	PD	VT	(3)/RARE//AN/
EE	VL	PT	VI	(3)/RARE//AN/
GE	C1	GE	GE	(1)/ITUTUARU/(2)(3)
GI	--	N1	N1	(1)/ITUTUARU/(2)(3)
GT	--	N1	N1	(1)/ITUTUARU/(2)(3)
L1	C1	L1	L1	(1)(2)(3)
L2	C1	L2	L2	(1)(2)(3)
L•	N3	AN	SS	(2)/ WA/(3)/ DE/(1)/RU/
=	GE	SS	SS	(2)/RU TO/(3)
=	GI	SS	SS	(2)/ITUTUARU/(3)
=	L2	SS	SS	(2)/ TAMENI/(3)
=	N1	SS	SS	(2)(3)
=	PA	SS	SS	(2)/ NI/(3)
=	PB	SS	SS	(2)(3)
=	PT	N1	N4	(2)/RARETA/(3)
=	VE	=	SS	(2)/ BESI/
AI	AS	N1	AI	(3)/ TO ONAJI/(1)
AI	B1	N1	N1	(1)/II/(2)(3)
AI	C1	N1	N1	(1)/KU/(2)(3)
AI	--	PI	N1	(1)/II/(2)(3)/IMASITA/
AI	PI	N1	N1	(1)/II/(2)/IMASITA/(3)
AI	PT	N1	N1	(1)/II/(2)/IMASITA/(3)
A•	SS	¥¥	SS	(2)(1)/RU// KA/
AN	AS	N4	AN	(3)/ TO ONAJI/(1)
AN	C1	AN	AN	(1)/NA/(2)(3)
AN	C4	AO	AO	(1)/NA/(2)(3)
AN	--	N1	N1	(1)/NA/(2)(3)
AN	--	PT	N1	(1)/NA/(2)(3)/IMASITA/
AN	PT	N1	N1	(1)/NA/(2)/IMASITA/(3)
AO	AS	N1	AO	(3)/ TO ONAJI/(1)
AO	C1	AO	AO	(1)/NO/(2)(3)
AO	--	AO	AO	(1)/NO/(2)(3)
AO	--	PT	N1	(1)/NO/(2)(3)/IMASITA/
AO	PT	N1	N1	(1)/NO/(2)/IMASITA/(3)
B1	AI	N1	N1	(1)(2)/II/(3)
B1	AO	N1	N1	(1)(2)/NO/(3)
B1	C1	N1	N1	(1)(2)(3)
B2	••	SS	SS	(1)(2)(3)
C2	SS	SS	SS	(2)(1)/ •/(3)
DT	••	N1	N4	(1)(3)
DT	B1	PI	DT	(1)(2)(3)/RARETA/
DT	DT	N1	N4	(1)(2)(3)
DT	GT	AO	N4	(1)(2)/ITUTUARU/(3)
DT	PD	N1	N4	(1)(2)/RARETA/(3)
DT	PT	N1	N4	(1)(2)/RARETA/(3)
DT	VT	N1	N4	(1)(2)/RU/(3)
EE	VH	PF	VT	(3)/IOE//AN/
EE	VH	PT	VT	(3)/IOE//AN/
EE	VL	GF	VT	(3)/ITUTU//AN/
EE	VL	GT	VT	(3)/ITUTU//AN/
EE	VL	PF	VI	(3)/RARE//AN/
F•	SS	¥¥	SS	(2)/ KA/
GI	C1	GI	GI	(1)/II/(2)(3)
GT	C1	GT	GT	(1)/II/(2)(3)
--	C1	--	C1	(1)(2)(3)
L1	C3	L1	L1	(1)(2)(3)
L•	N3	AI	SS	(2)/ WA/(3)/II/
L•	N3	AO	SS	(2)/ WA/(3)/ DE/(1)/RU/

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... A -- 3

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
N1	**	N1	N1	(1)/NO/(3)
N1	C1	N4	N1	(1)(2)(3)
N1	--	N1	N1	(1)(2)(3)
N1	PL	RS	SS	(1)/ WA/(3)/ KOTODE/(2)/IMASITA/
N3	AI	N1	N4	(1)(2)/II/(3)
N3	AO	N1	N4	(1)(2)/NO/(3)
N3	PE	RS	SS	(3)/ KOTOWA/(2)/IMASITA/
N3	PL	RS	SS	(1)/ WA/(3)/ KOTODE/(2)/IMASITA/
N3	VE	L?	SS	(3)/ KOTOWA/(2)
N3	VI	RS	SS	(1)/ WA/(3)/ KOTO O/(2)/RU/
N3	VL	RS	SS	(1)/ WA/(3)/ KOTODE/(2)/RU/
N4	B?	PE	SS	(2)/ ,/(1)/ WA/(3)/IMASITA/
N4	C1	N4	N4	(1)(2)(3)
N4	C4	N4	N4	(1)(2)(3)
N4	--	N4	N4	(1)(2)(3)
N4	N1	PE	N4	(2)/ GA/(3)(1)
N4	PL	RS	SS	(1)/ WA/(3)/ KOTODE/(2)/IMASITA/
N4	R1	PE	N4	(3)/IMASITA/(1)
N4	TH	PE	N4	(3)/IMASITA/(1)
N4	VE	VE	SS	(1)/ WA/(3)(2)
N4	VT	RS	SS	(1)/ WA/(3)/ KOTO O/(2)/RU/
P1	B?	N1	PA	(2)(3)(1)/ES/
P1	GI	L?	PA	(3)/ E/(2)/RU KOTO/(1)
P1	GT	L?	PA	(3)/ E/(2)/RU KOTO/(1)
P1	GT	N4	PA	(3)/ O/(2)/RU KOTO/(1)
P1	PI	N1	PA	(2)/RARETA/(3)(1)
P1	TH	N1	PA	(2)(3)(1)
P2	B1	N4	PB	(2)(3)(1)
P2	GI	L1	PB	(3)/ E/(2)/RU KOTO/(1)
P2	GT	L1	PB	(3)/ E/(2)/RU KOTO/(1)
P2	GT	N4	PB	(3)/ O/(2)/RU KOTO/(1)
P2	P1	N1	PB	(3)(2)(1)
P2	PT	N1	PB	(2)/RARETA/(3)(1)
P2	TH	RS	PB	(3)/ SORE/(1)
PA	C1	PA	PA	(1)(2)(3)
PA	**	PA	PA	(1)(2)(3)
PB	C1	PB	PB	(1)(2)(3)
PB	GT	L1	PB	(3)/ E/(2)/ITUTUARU/(1)
PD	N4	AN	PE	(2)/ O/(3)/ NI/(1)
PE	C1	PE	PE	(1)/IMASITA/(2)(3)
PF	EE	VF	VT	(3)/ANAKATADES/
PF	EE	VI	VI	(3)/ANAKATADES/
PH	EE	PD	VD	(3)/ANAKATADES/
PH	EE	PI	VI	(3)/ANAKATADES/
PH	EE	PT	VT	(3)/ANAKATADES/
PL	EE	GF	VT	(3)/ITUTU//ANAKATADES/
PL	EE	PF	VI	(3)/RARE//ANAKATADES/
PL	GL	PD	VT	(3)/RARE//TUTUAR/
PL	GL	PT	VI	(3)/RARE//TUTUAR/
PU	EE	AI	VI	(3)/KU/(1)/ANAKATADES/

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
N1	C1	N1	N1	(1)(2)(3)
N1	--	GT	N1	(1)(2)(3)/ITUTUARU/
N1	--	PT	N1	(1)(2)(3)/RARETA/
N1	VL	RS	SS	(1)/ WA/(3)/ KOTODE/(2)/RU/
N3	AN	N1	N4	(1)(2)/NA/(3)
N3	PE	L?	SS	(3)/ KOTOWA/(2)/IMASITA/
N3	PI	RS	SS	(1)/ WA/(3)/ KOTO O/(2)/IMASITA
N3	PT	RS	SS	(1)/ WA/(3)/ KOTO O/(2)/IMASITA
N3	VE	RS	SS	(3)/ KOTOWA/(2)
N3	VL	PB	SS	(1)/ WA/(3)/ DE/(2)/RU/
N3	VT	RS	SS	(1)/ WA/(3)/ KOTO O/(2)/RU/
N4	C1	N1	N4	(1)(2)(3)
N4	C4	N1	N4	(1)(2)(3)
N4	HH	VE	SS	(1)/ WA/(2)(3)
N4	N1	PE	N4	(2)/ GA/(3)/IMASITA/(1)
N4	N1	VE	N4	(2)/ GA/(3)(1)
N4	PT	RS	SS	(1)/ WA/(3)/ KOTO O/(2)/IMASITA
N4	R1	VE	N4	(3)(1)
N4	TH	VE	N4	(3)(1)
N4	VL	RS	SS	(1)/ WA/(3)/ KOTODE/(2)/RU/
P1	B1	N1	PA	(2)(3)(1)
P1	GD	N1	PA	(2)/ITUTUARU/(3)(1)
P1	GI	N1	PA	(2)/ITUTUARU/(3)(1)
P1	GT	N1	PA	(2)/ITUTUARU/(3)(1)
P1	N3	N1	PA	(2)(3)(1)
P1	PT	N1	PA	(2)/RARETA/(3)(1)
P2	B1	N1	PB	(2)(3)(1)
P2	B?	N4	PB	(2)(3)(1)
P2	GI	N1	PB	(2)/ITUTUARU/(3)(1)
P2	GT	N1	PB	(2)/ITUTUARU/(3)(1)
P2	N3	N1	PB	(2)(3)(1)
P2	PI	N1	PB	(2)/RARETA/(3)(1)
P2	TH	N1	PB	(2)(3)(1)
P4	PI	N1	PB	(2)/RARETA/(3)(1)
PA	C1	PB	PB	(1)/ NI(NO)/(2)(3)
PB	C1	PA	PA	(1)(2)(3)
PB	GI	L1	PB	(3)/ E/(2)/ITUTUARU/(1)
PD	N4	AI	PE	(2)/ O/(3)/KU/(1)
PD	N4	AO	PE	(2)/ O/(3)/ NI/(1)
PF	EE	VD	VD	(3)/ANAKATADES/
PF	EE	VH	VT	(3)/ANAKATADES/
PF	EE	VT	VT	(3)/ANAKATADES/
PH	EE	PF	VF	(3)/ANAKATADES/
PH	EE	PL	VL	(3)/ANAKATADES/
PL	EE	GD	VD	(3)/ITUTU//ANAKATADES/
PL	EE	GT	VT	(3)/ITUTU//ANAKATADES/
PL	EE	PT	VI	(3)/RARE//ANAKATADES/
PL	GL	PF	VI	(3)/RARE//TUTUAR/
PT	--	N1	N1	(1)/RARETA/(2)(3)
PU	EE	N4	VI	(3)/ DE/(1)/ANAKATADES/

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... A -- 3

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
PU	EE	PD	PD	(3)/RU/(1)/ANAKATADES/
PU	EE	VH	VH	(3)/RU/(1)/ANAKATADES/
PU	EE	VT	VT	(3)/RU/(1)/ANAKATADES/
QQ	..	SS	SS	(1)(2)(3)
R1	VE	¥¥	SS	(1)/ GA/(2)/ KA/
RA	P2	N4	AI	(3)(2)(1)
RS	C3	RS	RS	(1)(2)(3)
SS	AS	SS	SS	(3)/ NI TURETE/(1)
SS	B2	¥¥	SS	(2)(1)/ KA/
SS	C1	SS	SS	(1)(2)(3)
SS	C3	SS	SS	(1)(2)(3)
SS	¥¥	=	SS	(1)/ KA/
SS	..	GE	SS	(3)/ITUTU/(2)(1)
TH	PA	SS	RS	(2)/ NI/(3)
TO	AI	N1	L1	(2)/II/(3)
TO	GI	N1	L1	(2)/ITUTUARU/(3)
TO	N1	PE	L1	(3)/RARETA/(2)
TO	N4	PE	L1	(3)/RARETA/(2)
VA	AO	VT	VT	(2)/ NI/(3)/RU/(1)
VA	EE	VF	VF	(3)(1)/AN/
VA	EE	VI	VI	(3)(1)/AN/
VA	EE	VT	VT	(3)(1)/AN/
VA	VH	PF	VT	(3)/IOE/(1)
VA	VH	PT	VT	(3)/IOE/(1)
VA	VL	GF	VT	(3)/ITUTU/(2)(1)
VA	VL	GT	VT	(3)/ITUTU/(2)(1)
VA	VL	PF	VI	(3)/RARERU/
VA	VL	PT	VI	(3)/RARE/(1)
VD	N4	AI	VE	(2)/ O/(3)/KU/(1)/RU/
VD	N4	AO	VE	(2)/ O/(3)/ NI/(1)/RU/
VE	C3	VE	VE	(1)(2)(3)
VF	EE	VD	VI	(3)/AN/
VF	EE	VH	VH	(3)/AN/
VF	EE	VT	VT	(3)/AN/
VH	DT	PF	PI	(2)(3)
VH	EE	PF	VF	(3)/ANAKATADES/
VH	EE	PI	VI	(3)/ANAKATADES/
VH	EE	PT	VT	(3)/ANAKATADES/
VL	B1	EE	VL	(2)/AN/(1)
VL	EE	AI	VE	(3)/KU//ANAI/
VL	EE	GF	VT	(3)/ITUTU//AN/
VL	EE	N4	VE	(3)/ DE/(1)/AN//RU/
VL	EE	PF	VI	(3)/RARE//AN/
VL	GL	PD	VD	(3)/RARE//AN/
VL	GL	PT	VI	(3)/RARE//TUTUAR/
VT	C1	VT	VT	(1)/RU/(2)(3)
WH	PL	N3	SS	(3)/ WA/(1)/ DE/(2)/IMASITA/
WS	C1	WS	WS	(1)(2)(3)
..	AN	C3	C4	(1)(2)(3)
..	GE	C3	C4	(1)(2)(3)

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
PU	EE	VF	VF	(3)/RU/(1)/ANAKATADES/
PU	EE	VI	VI	(3)/RU/(1)/ANAKATADES/
PU	SS	¥¥	SS	(2)(1)/IMASITA// KA/
R1	PE	¥¥	SS	(1)/ GA/(2)/IMASITA// KA/
RA	P2	N1	AI	(3)(2)(1)
RS	C1	RS	RS	(1)(2)(3)
**	C3	SS	**	(1)(2)(3)
SS	B1	¥¥	SS	(2)/ ,/(1)/ KA/
SS	C1	PE	SS	(1)(2)(3)/IMASITA/
SS	C1	VE	SS	(1)(2)(3)
SS	N4	¥¥	SS	(2)/ ,/(1)/ KA/
SS	..	B2	SS	(3)(1)
SS	..	SS	SS	(1)(2)(3)
TH	PB	SS	RS	(2)(3)
TO	DT	N1	L1	(2)(3)
TO	GT	N4	L1	(2)/ITUTUARU/(3)/ES/
TO	N1	VT	L1	(2)(3)
TO	VF	N3	L2	(3)/ O/(2)/RU/
VA	EE	VD	VD	(3)(1)/AN/
VA	EE	VH	VH	(3)/ NODE/(1)/AN// NODE/
VA	EE	VL	VL	(3)(1)/AN/
VA	SS	¥¥	SS	(2)(1)/RU// KA/
VA	VH	PI	VI	(3)/IOE/(1)
VA	VL	GD	VD	(3)/ITUTU/(2)(1)
VA	VL	GI	VI	(3)/ITUTU/(2)(1)
VA	VL	PD	VT	(3)/RARE/(1)
VA	VL	PF	VT	(3)/RARE/(1)
VA	VL	VT	VI	(3)/RARE/(1)
VD	N4	AN	VE	(2)/ O/(3)/ NI/(1)/RU/
VE	C1	VE	VE	(1)(2)(3)
VE	C4	VE	VE	(1)(2)(3)
VF	EE	VF	VF	(3)/AN/
VF	EE	VI	VI	(3)/AN/
VF	SS	¥¥	SS	(2)/ KA/
VH	EE	PD	VD	(3)/ANAKATADES/
VH	EE	PH	VT	(3)/ANAKATADES/
VH	EE	PL	VL	(3)/ANAKATADES/
VH	EE	VL	VL	(3)/ANAKATADES/
VL	B1	GT	VT	(2)(3)/ITUTU/(1)
VL	EE	GD	VD	(3)/ITUTU//AN/
VL	EE	GT	VT	(3)/ITUTU//AN/
VL	EE	PD	VT	(3)/RARE//AN/
VL	EE	PT	VI	(3)/RARE//AN/
VL	GL	PF	VI	(3)/RARE//TUTUAR/
VL	RA	PE	VE	(2)/KU/(3)/RARERU/
VT	N4	L2	VE	(2)/ GA/(3)/ KOTO O/(1)/RU/
WH	VL	N3	SS	(3)/ WA/(1)/ DE/(2)/RU/
XX	VT	=	SS	(1)(2)/RU BESI/
..	B1	..	B2	(2)
..	GE	..	B2	(1)/RU TO/(2)(3)

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... A -- 3

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
,,	L2	SS	SS	(1)(2)/ TAMENI/(3)
,,	N1	C4	C4	(1)(2)(3)
,,	N4	C3	C4	(1)(2)(3)
,,	N4	))	))	(1)(2)(3)
,,	PA	SS	SS	(1)(2)/ NI/(3)
,,	PB	,,	PB	(2)
,,	RS	,,	RS	(2)
,,	VE	C4	C4	(1)(2)(3)
((	AI	))	N1	(1)(2)/II/(3)
((	AO	))	N1	(1)(2)/NO/(3)
((	N4	))	N1	(1)(2)(3)
((	PB	))	PB	(1)(2)(3)
((	SS	))	N4	(1)(2)(3)
((	VI	))	N1	(1)(2)/RU/(3)
,,	N1	C3	C4	(1)(2)(3)
,,	N1	))	))	(1)(2)(3)
,,	N4	C4	C4	(1)(2)(3)
,,	PA	=	B2	(2)/ NI/(1)
,,	PA	,,	PA	(2)
,,	PT	,,	B2	(2)/RARETA/
,,	VE	C3	C4	(1)(2)(3)
,,	WS	,,	B2	(1)(2)(3)
((	AN	))	N1	(1)(2)/NA/(3)
((	N1	))	N1	(1)(2)(3)
((	PA	))	PA	(1)(2)(3)
((	RS	))	RS	(1)(2)(3)
((	VE	))	VE	(1)(2)(3)
((	VT	))	N1	(1)(2)/RU/(3)

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... A -- 2

PATTERN (1)	PATTERN (2)	SUBSTITUTE	CORRESPONDING-JAPANESE
=	HH	UU	(2)
=	PA	B2	(2)/ NI/
=	PF	FP	(2)
=	PU	AP	(2)
=	SS	SS	(2)
=	VF	F.	(2)
=	VL	L.	(2)
AI	AN	N1	(1)/II/(2)
AI	N1	N1	(1)/II/(2)
AN	AN	N1	(1)/NA/(2)
AO	AN	AN	(1)/NO/(2)
AO	L1	L1	(2)
AO	PB	N4	(2)(1)
AS	AI	AI	(2)
AS	PA	B2	(2)/ TO ONAJI// NI/
B1	AI	AI	(1)(2)
B1	AO	AO	(1)(2)
B1	GI	GI	(1)(2)
B1	N1	N1	(1)(2)
B1	RA	RA	(1)(2)
B1	VE	VE	(1)(2)
B2	SS	SS	(1)(2)
C1	PA	PA	(1)(2)
C2	SS	B2	(2)(1)
DT	AI	N4	(1)(2)/II// MONO/
DT	AO	N4	(1)(2)
DT	N1	N4	(1)(2)
DT	PA	PA	(2)
DT	RA	N4	(1)(2)/ MONO/
DT	VI	N4	(1)(2)/RU KOTO/
DY	N1	B1	(1)(2)/ NI/
FP	SS	SS	(2)/TA/
GH	B1	GH	(1)(2)
GI	AI	GE	(2)/KU/(1)
GI	AO	GE	(2)/ NI/(1)
GL	AN	GE	(2)/ DE/(1)
GL	PT	GI	(2)/RARE/
GT	AO	GE	(2)/ O/(1)
GT	RS	GE	(2)/ KOTO O/(1)
--	N1	N1	(1)(2)
L1	PB	L1	(2)(1)
L1	WS	L1	(2)(1)
N1	N1	N1	(1)(2)
N3	AO	N4	(1)(2)
N4	GE	N4	(2)/ITUTUARU/(1)
N4	PA	N4	(2)/ NI(NO)/(1)
N4	RS	N4	(2)(1)
P1	=	B1	(1)
P1	AN	PA	(2)(1)
P1	GE	PA	(2)/RU/(1)
=	N1	N4	(2)
=	PB	B2	(2)
=	PL	LP	(2)
=	R1	WH	(2)
=	VA	A.	(2)
=	VH	H.	(2)
AI	AI	AI	(1)/II/(2)
AI	AO	N1	(1)/II/(2)
A.	SS	SS	(2)(1)/RU/
AN	N1	N1	(1)/NA/(2)
AO	AO	N1	(1)/NO/(2)
AO	N1	N1	(1)/NO/(2)
AP	SS	SS	(2)(1)/TA/
AS	L1	PB	
AS	PE	B2	(2)/RARETA// TO ONAJI/
B1	AN	AN	(1)(2)
B1	B1	B1	(1)(2)
B1	GT	GT	(1)(2)
B1	PE	PE	(1)(2)
B1	SS	SS	(1)(2)
B1	WS	B2	(1)(2)
B2	VE	VE	(1)(2)
C1	SS	SS	(1)(2)
DT	=	B1	(2)/ NI/
DT	AN	N4	(1)(2)
DT	B1	N4	(1)(2)/ MONO/
DT	N4	N4	(1)(2)
DT	PB	N4	(2)(1)/ MONO/
DT	RS	RS	(2)
DT	VT	N4	(1)(2)/RU KOTO/
F.	SS	SS	(2)
GE	B1	GE	(2)(1)
GI	=	GE	(1)
GI	AN	GE	(2)/ NI/(1)
GL	AI	GE	(2)/KU/(1)
GL	AO	GE	(2)/ DE/(1)
GT	AN	GE	(2)/ O/(1)
GT	N3	GE	(2)/ O/(1)
HH	=	B1	(1)
L1	PA	L1	(2)/NO/(1)
L1	RS	L1	(2)(1)
N1	**	N1	(1)/NO// MONO/
N3	AN	N4	(1)(2)
N3	N1	N4	(1)(2)
N4	L1	N4	(2)/ E NO/(1)
N4	PB	N4	(2)(1)
N4	WS	N4	(2)(1)
P1	AI	PA	(2)/I MONO/(1)
P1	AO	PA	(2)(1)
P1	N3	PA	(2)(1)

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... A -- 2

PATTERN (1)	(2)	SUBSTI-TUTE	CORRESPONDING-JAPANESE
P1	N4	PA	(2)(1)
P1	R1	R1	(2)
P2	AI	PB	(2)/I MONO/(1)
P2	AO	PB	(2)(1)
P2	N3	PB	(2)(1)
P2	PB	PB	(2)(1)
P4	GE	PB	(2)/RU KOTO/(1)
PA	PA	PA	(2)/ NI(NO)/(1)
PA	RS	PA	(2)(1)
PB	L1	PB	(2)/ E NO/(1)
PB	PB	PB	(2)(1)
PB	WS	PB	(2)(1)
PH	PF	PT	(2)
PH	PI	PI	(2)
PH	PT	PT	(2)
PI	B1	PI	(2)(1)
PI	GT	PE	(2)/I NI/(1)
PL	AI	VE	(2)/KATTA/
PL	AO	VE	(2)/ DE/(1)/IMASITA/
PL	EE	VL	(1)/ANAKATADES/
PL	GF	PT	(2)/ITUTU/(1)
PL	GT	PT	(2)/ITUTU/(1)
PL	PE	PE	(2)/RARE/
PL	PI	PI	(2)/RARE/
PT	B1	PT	(2)(1)
PT	WS	PE	(2)/ KOTO O/(1)
PU	VE	VE	(2)(1)/IMASITA/
PU	VH	PT	(2)(1)
PU	VL	PL	(2)(1)
QQ	=	SS	(1)
QQ	GT	N1	(2)/ BESI// KOTO/
QQ	SS	SS	(1)(2)
R3	SS	SS	(1)(2)
RA	AI	AI	(1)(2)
RA	AO	AO	(1)(2)/ES/
RA	N1	N1	(1)/II/(2)
SS	=	SS	(1)
SS	B2	SS	(2)/ ,/(1)
SS	L2	SS	(2)/ KOTOWA/(1)
SS	PA	SS	(2)/ NI/(1)
SS	¥¥	SS	(1)/ KA/
TO	AO	L1	(2)
TO	N1	L1	(2)
TO	N4	L1	(2)
VA	EE	VA	(1)(2)
VA	VE	VE	(2)(1)/RU/
VA	VH	VT	(2)(1)
VA	VL	VL	(2)(1)
VD	RS	VE	(2)/ O/(1)/RU/
VE	AN	VE	(2)/NA/(1)

PATTERN (1)	(2)	SUBSTI-TUTE	CORRESPONDING-JAPANESE
P1	PA	PA	(2)(1)
P2	=	B1	(1)
P2	AN	PB	(2)(1)
P2	GE	PB	(2)/RU KOTO/(1)
P2	N4	PB	(2)(1)
P2	R1	R1	(2)
PA	L1	PA	(2)/ E NO/(1)
PA	PB	PA	(2)(1)
PA	WS	PA	(2)(1)
PB	PA	PB	(2)/ NI(NO)/(1)
PB	RS	PB	(2)(1)
PF	VE	VE	(2)
PH	PH	PT	(2)
PH	PL	PL	(2)
PI	AI	VE	(2)/KU/(1)/IMASITA/
PI	GI	PE	(2)/I NI/(1)
PI	WS	PE	(2)/ KOTO O/(1)
PL	AN	VE	(2)/ DE/(1)/IMASITA/
PL	B1	PL	(2)(1)
PL	GD	PD	(2)/ITUTU/(1)
PL	GI	PI	(2)/ITUTU/(1)
PL	PD	PT	(2)/RARE/
PL	PE	PI	(2)/RARE/
PL	PT	PI	(2)/RARE/
PT	RS	PE	(2)/ KOTO O/(1)
PU	VD	PD	(1)(2)
PU	VF	PT	(2)(1)
PU	VI	VI	(2)(1)
PU	VT	VT	(2)(1)
QQ	B1	SS	(2)(1)
QQ	N1	N4	(1)(2)
R1	SS	RS	(2)
RA	=	B1	(1)/KU/
RA	AN	AN	(1)/KU/(2)
RA	B1	B1	(1)(2)
RA	PB	AI	(2)(1)
SS	B1	SS	(2)(1)
SS	HH	SS	(2)(1)
SS	N5	SS	(1)(2)
SS	PB	SS	(2)(1)
TH	SS	RS	(2)
TO	GE	PB	(2)/ E NO/
TO	N3	L1	(2)(1)
TO	VE	L2	(2)
VA	VD	VD	(2)(1)
VA	VF	VT	(2)(1)
VA	VI	VI	(2)(1)
VA	VT	VT	(2)(1)
VE	AI	VE	(2)/II/(1)
VE	AO	VE	(2)/NO/(1)

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... A -- 2

PATTERN (1)	(2)	SUBSTITUTE	CORRESPONDING-JAPANESE
VE	B1	VE	(2)(1)
VE	L2	VE	(2)/ TAMENI/(1)
VF	PA	VE	(2)/ NI/(1)/RU/
VH	=	VE	(1)/RU/
VH	PE	VE	(2)/IMASITA/
VH	PH	PT	(2)
VH	PL	PL	(2)
VH	VI	PI	(2)
VI	=	VE	(1)/RU/
VI	B2	VI	(2)(1)
VI	WS	VE	(2)/ KOTO O/(1)/RU/
VL	AI	VE	(2)/II/
VL	AO	VE	(2)/ DE/(1)/RU/
VL	B2	VL	(2)(1)
VL	GD	VD	(2)/ITUTU/(1)
VL	GI	VI	(2)/ITUTU/(1)
VL	L2	VE	(2)/ YOTEDE/(1)/RU/
VL	PD	VT	(2)/RARE/
VL	PI	VI	(2)/RARE/
VL	QQ	VL	(1)/AN/
VT	B1	VT	(2)(1)
VT	GE	VE	(2)/RU MONO/(1)/RU/
VT	GT	VE	(2)/RU MONO/(1)/RU/
VT	SS	VE	(2)/ KOTO O/(1)/RU/
W1	SS	SS	(1)(2)
W2	AN	WA	(1)(2)
W2	SS	N4	(1)(2)/ KA/
WH	SS	SS	(1)/ O/(2)
WN	SS	SS	(2)/ TOKI/
..	PE	PE	(2)
..	WS	B2	(2)(1)

PATTERN (1)	(2)	SUBSTITUTE	CORRESPONDING-JAPANESE
VE	GE	VE	(2)/ITUTU/(1)
VF	EE	VF	(1)/AN/
VF	VE	VE	(2)
VH	PD	PD	(2)
VH	PF	PT	(2)
VH	PI	PI	(2)
VH	PT	PT	(2)
VH	VT	PT	(2)
VI	B1	VI	(2)(1)
VI	GI	VE	(2)/I NI/(1)/RU/
VL	=	VE	(1)/RU/
VL	AN	VE	(2)/ DE/(1)/RU/
VL	B1	VL	(2)(1)
VL	EE	VL	(1)/AN/
VL	GF	VT	(2)/ITUTU/(1)
VL	GT	VT	(2)/ITUTU/(1)
VL	PB	VE	(2)(1)
VL	PF	VI	(2)/RARE/
VL	PT	VI	(2)/RARE/
VT	=	VE	(1)/RU/
VT	B2	VT	(2)(1)
VT	GE	VE	(2)/RU MONO/(1)/RU/
VT	RS	VE	(2)/ KOTO O/(1)/RU/
VT	WS	VE	(2)/ KOTO O/(1)/RU/
W2	AI	WI	(1)(2)
W2	L2	N4	(2)/ HOHO/
WA	SS	SS	(1)/NA/(2)
WI	SS	SS	(1)/KU/(2)
..	C1	C3	(1)(2)
..	RS	RS	(2)
((	))	N1	(1)(2)

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ••• B -- 3

PATTERN (1)	PATTERN (2)	PATTERN (3)	SUBSTITUTE	CORRESPONDING-JAPANESE
=	GE	••	PA	(2)/RU TO/
=	L2	VE	SS	(2)/ KOTOWA/(3)
=	RS	PE	SS	(2)/ KOTOWA/(3)/IMASITA/
=	VI	L1	SS	(3)/ E/(2)/RU BESI/
AS	N4	VE	B2	(2)/ GA/(3)/ NI TURETE/
AS	TH	PE	RS	(3)/RARETA// MONO/
C2	N4	VE	B2	(2)/ GA/(3)(1)
HH	••	N4	SS	(3)/ DE ARU/
HH	VL	N4	SS	(3)/ GA/(2)/RU/
L1	TH	VE	L1	(3)(1)
L•	N3	N4	SS	(2)/ WA/(3)/ DE/(1)/RU/
L•	N4	GE	SS	(2)/ WA/(3)/ITUTUARU/
L•	TH	N4	SS	(2)/ WA/(3)/ DE/(1)/RU/
N3	••	GI	SS	(1)/ WA/(3)/ITUTUARU/
N3	VF	=	SS	(1)/ WA/
N4	••	N4	SS	(1)/ WA/(3)/ DE ARU/
N4	B2	PE	SS	(2)/ •/(1)/ WA/(3)/IMASITA/
N4	GE	VE	SS	(2)/ITUTUARU/(1)/ WA/(3)
N4	N4	PE	N4	(2)/ GA/(3)/IMASITA/(1)
N4	N5	VE	N4	(2)/ GA/(3)(1)
N4	PE	VE	SS	(2)/RARETA/(1)/ WA/(3)
N4	PT	VE	SS	(2)/RARETA/(1)/ WA/(3)
N4	R1	VE	N4	(3)(1)
N4	TH	VE	N4	(3)(1)
N4	VA	=	SS	(1)/ WA/
N4	VF	=	SS	(1)/ WA/
N4	••	SS	SS	(1)(2)(3)
P4	N4	PE	B2	(2)/ GA/(3)/IMASITA// NODE/
PB	R1	VE	PR	(3)(1)
PD	N4	L1	PE	(3)/ NI/(2)/ O/(1)
PH	QQ	N4	PH	(3)/ O/(1)/ANAKATADES//RU/
PT	N4	PI	VE	(3)/RARETA/(2)/ O/(1)/IMASITA/
R1	AO	PE	RS	(2)/ NI/(3)/IMASITA/
R1	N4	PE	RS	(2)/ GA/(3)/IMASITA/
R1	N4	PH	RS	(2)/ GA/(3)/IMASITA/
R1	N4	VD	RS	(2)/ GA/(3)/RU/
R1	N4	VF	RS	(2)/ GA/(3)/RU/
R1	N4	VT	RS	(2)/ GA/(3)/RU/
R1	PA	VE	RS	(2)/ NI/(3)
R1	PB	VE	RS	(2)(3)
R1	VL	AI	RS	(3)/II/
R1	VL	TH	SS	(3)/ WA/(1)/ DE/(2)/RU/
R2	N4	VE	RS	(1)(2)/ GA/(3)
R3	N4	PT	RS	(2)/ GA/(3)/IMASITA/
R3	N4	VT	RS	(2)/ GA/(3)/RU/
RA	P2	N4	AI	(3)(2)(1)
SS	••	GE	SS	(1)(2)(3)/RU TO/
TH	N4	VE	RS	(2)/ GA/(3)
TH	VT	N4	RS	(3)/ O/(2)/RU/
UU	VL	N4	SS	(3)/ GA/(2)/RU/
=	L2	PE	SS	(2)/ KOTOWA/(3)/IMASITA/
=	N4	••	N4	(2)
=	RS	VE	SS	(2)/ KOTOWA/(3)
=	XX	VE	SS	(2)(3)/ BESI/
AS	N4	VT	SS	(2)/ GA/(3)/RU// NI TURETE/
C2	N4	PE	B2	(2)/ GA/(3)/IMASITA/(1)
GD	N4	N4	GE	(2)/ NI/(3)/ O/(1)
HH	PL	N4	SS	(3)/ GA/(2)/IMASITA/
H•	N4	N4	SS	(2)/ WA/(3)/ O/(1)/RU/
L•	HH	N4	SS	(3)/ GA/(1)/RU/
L•	N4	AI	SS	(2)/ WA/(3)/II/
L•	N4	N4	SS	(2)/ WA/(3)/ DE/(1)/RU/
N3	••	AI	SS	(1)/ WA/(3)/II/
N3	••	N4	SS	(1)/ WA/(3)/ DE ARU/
N4	••	GE	SS	(1)/ WA/(3)/ITUTUARU/
N4	AS	N4	N4	(3)/ TO ONAJI/(1)
N4	B2	VE	SS	(2)/ •/(1)/ WA/(3)
N4	GT	VE	SS	(2)/ITUTUARU/(1)/ WA/(3)
N4	N4	PI	N4	(2)/ GA/(3)/IMASITA/(1)
N4	PE	PI	SS	(2)/RARETA/(1)/ WA/(3)/IMASITA/
N4	PF	=	SS	(1)/ WA/
N4	R1	PE	N4	(3)/IMASITA/(1)
N4	TH	PE	N4	(3)/IMASITA/(1)
N4	TH	VT	N4	(3)/RU/(1)
N4	VE	PE	SS	(1)/ WA/(3)/RARETA/(2)
N4	VI	RS	SS	(3)(1)/ WA/(2)/RU/
N5	AN	VE	SS	(2)/NA/(1)/ WA/(3)
P4	N4	VE	B2	(2)/ GA/(3)/ NODE/
PD	N3	L1	PE	(2)/ O/(3)/ NI/(1)
PD	N4	N4	PE	(2)/ NI/(3)/ O/(1)
PT	N4	PE	VE	(3)/RARETA/(2)/ O/(1)/IMASITA/
PT	N4	PT	VE	(3)/RARETA/(2)/ O/(1)/IMASITA/
R1	N4	PD	RS	(2)/ GA/(3)/IMASITA/
R1	N4	PF	RS	(2)/ GA/(3)/IMASITA/
R1	N4	PT	RS	(2)/ GA/(3)/IMASITA/
R1	N4	VE	RS	(2)/ GA/(3)
R1	N4	VH	RS	(2)/ GA/(3)/RU/
R1	PA	PE	RS	(2)/ NI/(3)/IMASITA/
R1	PB	PE	RS	(2)(3)/IMASITA/
R1	PT	N4	RS	(3)/ O/(2)/IMASITA/
R1	VL	N4	RS	(3)/ DE/(2)/RU/
R1	VT	N4	RS	(3)/ O/(2)/RU/
R3	N4	PI	RS	(2)/ GA/(3)/IMASITA/
R3	N4	VI	RS	(2)/ GA/(3)
RA	P2	DT	AI	(3)(2)(1)
**	N4	••	N4	(2)
TH	N4	PE	RS	(2)/ GA/(3)/IMASITA/
TH	PT	N4	RS	(3)/ O/(2)/IMASITA/
UU	PL	N4	SS	(3)/ GA/(2)/IMASITA/
VA	N4	VI	SS	(2)/ WA/(3)(1)/RU/

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... B -- 3

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
VA	N4	VT	SS	(2)/ WA/(3)(1)/RU/
VD	N4	AN	VE	(2)/ GA/(3)/ DE ARU/(1)/RU/
VD	N4	N4	VE	(2)/ NI/(3)/ O/(1)/RU/
VH	QQ	N4	VE	(3)/ O/(1)/ANAI/
VL	N4	AI	SS	(2)/ WA/(3)/II/
VL	N4	GI	SS	(2)/ WA/(3)/ITUTU/(1)/RU/
VL	N4	N4	SS	(2)/ WA/(3)/ DE/(1)/RU/
VL	N4	PT	VE	(3)/IMASITA/(2)/ DE/(1)/RU/
VT	N4	AI	VE	(3)/II/(2)/ O/(1)/RU/
VT	N4	AO	VE	(2)/ O/(3)/ NI/(1)/RU/
VT	N4	PE	VE	(3)/RARETA/(2)/ O/(1)/RU/
VT	N4	PT	VE	(3)/RARETA/(2)/ O/(1)/RU/
VT	W1	SS	VE	(2)(3)/ KA O/(1)/RU/
W1	N4	PE	WS	(2)/ GA/(3)/IMASITA/
W1	N4	VE	WS	(2)/ GA/(3)
W1	VF	SS	SS	(1)(3)
WH	N4	PE	N4	(2)/ GA/(3)/IMASITA// KOTO/
WH	PL	N3	SS	(3)/ WA/(1)/ DE/(2)/IMASITA/
WH	PL	PA	SS	(3)/ WA/(1)/ DE/(2)/IMASITA/
WH	VH	N4	SS	(3)/ WA/(1)/ O/(2)/RU/
WH	VL	N4	SS	(3)/ WA/(1)/ DE/(2)/RU/
WI	VL	N4	SS	(3)/ WA/(1)/KU/(2)/RU/
WN	N4	PE	B2	(2)/ GA/(3)/IMASITA// TOKI/
WN	VF	SS	SS	(1)(3)
XX	N4	VE	SS	(2)/ GA/(3)/ TOSEYO/
XX	VT	N4	SS	(1)(3)/ O/(2)/ BESI/
..	PE	..	B2	(2)/RARETA/

PATTERN (1)	(2)	(3)	SUBSTITUTE	CORRESPONDING-JAPANESE
VD	N3	L1	VE	(2)/ O/(3)/ NI/(1)/RU/
VD	N4	L1	VE	(3)/ NI/(2)/ O/(1)/RU/
VD	N4	VE	VE	(2)/ GA/(3)/ KOTO O/(1)/RU/
VI	N4	VE	VE	(3)(2)/ O/(1)/RU/
VL	N4	GE	SS	(2)/ WA/(3)/ITUTU/(1)/RU/
VL	N4	GT	SS	(2)/ WA/(3)/ITUTU/(1)/RU/
VL	N4	PE	VE	(3)/IMASITA/(2)/ DE/(1)/RU/
VL	QQ	N4	VE	(3)/ GA/(1)/AN//RU/
VT	N4	AN	VE	(2)/ O/(3)/ NI/(1)/RU/
VT	N4	N4	VE	(2)/ NI/(3)/ O/(1)/RU/
VT	N4	PI	VE	(3)/RARETA/(2)/ O/(1)/RU/
VT	N4	VI	VE	(2)/ GA/(3)/RU MONO/(1)/RU/
W1	..	N4	SS	(3)/ WA/(1)/ DE ARU/
W1	N4	PI	WS	(2)/ GA/(3)/IMASITA/
W1	N4	VI	WS	(2)/ GA/(3)/RU/
WH	..	N4	SS	(3)/ WA/(1)/ DE ARU/
WH	N4	VE	N4	(2)/ GA/(3)/ KOTO/
WH	PL	N4	SS	(3)/ WA/(1)/ DE/(2)/IMASITA/
WH	VF	SS	SS	(1)/ O/(3)
WH	VL	N3	SS	(3)/ WA/(1)/ DE/(2)/RU/
WH	VL	PA	SS	(3)/ WA/(1)/ DE/(2)/RU/
WN	N3	VI	SS	(2)/ GA/(3)/RU// TOKI/
WN	N4	VE	B2	(2)/ GA/(3)/ TOKI/
XX	..	VE	SS	(1)(3)/ BESI/
XX	N4	VI	SS	(2)/ O/(3)/ASIMEYO/
..	N4	=	B2	(2)
..	PE	..	PB	(2)/RARETA/

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... B -- 2

PATTERN (1)	PATTERN (2)	SUBSTITUTE	CORRESPONDING-JAPANESE
=	VE	SS	(2)/ BESI/
AI	L1	AI	(2)/ E NO/(1)
AI	PB	AI	(2)(1)
AN	PA	N4	(2)/ NI(NO)/(1)
AO	VE	SS	(1)/ WA/(2)
AS	AN	AN	(2)
AS	N4	PB	(2)/ TO ONAJI/
B1	PD	PD	(1)(2)
B1	PH	PH	(1)(2)
B1	PL	PL	(1)(2)
B1	PU	PU	(1)(2)
B1	VD	VD	(1)(2)
B1	VH	VH	(1)(2)
B1	VL	VL	(1)(2)
EE	N4	N4	(1)(2)
GE	L1	GE	(2)/ E/(1)
GE	PA	GE	(2)/ NI/(1)
GH	N4	GE	(2)/ O/(1)
GI	N4	GE	(2)/ NI/(1)
GI	PB	GE	(2)(1)
GT	L1	GE	(2)/ E/(1)
GT	N4	GE	(2)/ O/(1)
GT	PB	GE	(2)(1)
--	VE	SS	/ES//ES//ES/
L1	PE	L1	(2)/RARETA/(1)
L2	VE	SS	(1)/ KOTOWA/(2)
N3	VE	SS	(1)/ WA/(2)
N4	L2	N4	(2)/ TAMENO/(1)
N4	P?	N4	(2)(1)
N4	RS	N4	(2)(1)
N4	VH	SS	(1)/ WA/(2)/RU/
N4	VI	SS	(1)/ WA/(2)/RU/
N5	PE	SS	(1)/ WA/(2)/IMASITA/
P1	B1	B2	(2)
P4	N4	PB	(2)(1)
PA	PE	PA	(2)/RARETA/(1)
PD	AN	PE	(2)/ NI/(1)
PD	L1	PE	(2)/ E/(1)
PD	N3	PE	(2)/ O/(1)
PD	PA	PE	(2)/ NI/(1)
PE	L1	PE	(2)/ E/(1)
PE	PA	PE	(2)/ NI/(1)
PH	N3	VE	(2)/ O/(1)/IMASITA/
PH	PA	VE	(2)/ NI/(1)/IMASITA/
PI	=	PE	(1)
PI	L2	PE	(2)/ TAMENI/(1)
PI	N4	PE	(2)/ NI/(1)
PI	PB	PE	(2)(1)
PL	N4	PE	(2)/ DE/(1)
PT	=	PE	(1)
=	VT	SS	(2)/RU// BESI/
AI	PA	AI	(2)/ NI/(1)
AN	=	N4	(1)
AO	=	N4	(1)
AS	AI	AI	(2)
AS	AO	AO	(2)
B1	N4	N4	(1)(2)
B1	PF	PF	(1)(2)
B1	PI	PI	(1)(2)
B1	PT	PT	(1)(2)
B1	VA	VA	(1)(2)
B1	VF	VF	(1)(2)
B1	VI	VI	(1)(2)
B1	VT	VT	(1)(2)
GD	N4	GE	(2)/ O/(1)
GE	L2	GE	(2)/ TAMENI/(1)
GE	PB	GE	(2)(1)
GI	L1	GE	(2)/ E/(1)
GI	PA	GE	(2)/ NI/(1)
GL	N4	GE	(2)/ DE/(1)
GT	N3	GE	(2)/ O/(1)
GT	PA	GE	(2)/ NI/(1)
HH	VE	SS	(2)
L1	GE	L1	(2)/ITUTUARU/(1)
L2	PE	SS	(1)/ KOTOWA/(2)/IMASITA/
N3	PE	SS	(1)/ WA/(2)/IMASITA/
N4	GE	N4	(2)/ITUTUARU/(1)
N4	PA	N4	(2)/ NI(NO)/(1)
N4	PI	SS	(1)/ WA/(2)/RU/
N4	VH	SS	(1)/ WA/(2)/RU/
N4	VI	SS	(1)/ WA/(2)/RU/
N4	WS	N4	(2)(1)
N5	VE	SS	(1)/ WA/(2)
P4	N3	PB	(2)(1)
P4	SS	B2	(2)/ NODE/
PB	PE	PB	(2)/RARETA/(1)
PD	AO	PE	(2)/ NI/(1)
PD	L2	PE	(2)/ KOTO O/(1)
PD	N4	PE	(2)/ O/(1)
PD	PB	PE	(2)(1)
PE	L2	PE	(2)/ TAMENI/(1)
PE	PB	PE	(2)(1)
PH	N4	VE	(2)/ O/(1)/IMASITA/
PH	PB	VE	(2)(1)/IMASITA/
PI	L1	PE	(2)/ E/(1)
PI	N3	PE	(2)/ NI/(1)
PI	PA	PE	(2)/ NI/(1)
PL	=	PE	(1)
PL	PA	VE	(2)/ DE/(1)/IMASITA/
PT	AN	PE	(2)/ O/(1)

Table 4.5.1 SYNTACTIC PATTERNS FOR MECHANICAL TRANSLATION ... B -- 2

PATTERN (1)	(2)	SUBSTITUTE	CORRESPONDING-JAPANESE
PT	AO	PE	(2)/ O/(1)
PT	L2	PE	(2)/ TAMENI/(1)
PT	N4	PE	(2)/ O/(1)
PT	PB	PE	(2)(1)
R1	PI	RS	(2)/IMASITA/
R1	VE	RS	(2)
R1	VT	RS	(2)/RU/
RA	PB	AI	(2)(1)
**	SS	**	(1)(2)/ • /
TO	PE	L2	(2)/RU/
TO	VH	L2	(2)/RU/
TO	VT	L2	(2)/RU/
VD	AN	VE	(2)/ NI/(1)/RU/
VD	L1	VE	(2)/ E/(1)/RU/
VD	N3	VE	(2)/ O/(1)/RU/
VD	PA	VE	(2)/ NI/(1)/RU/
VE	L1	VE	(2)/ E/(1)
VE	PA	VE	(2)/ NI/(1)
VF	SS	SS	(2)
VH	N4	VE	(2)/ O/(1)/RU/
VH	PB	VE	(2)(1)/RU/
VI	AN	VE	(2)/ NI/(1)/RU/
VI	GE	VE	(2)/ITUTU/(1)/RU/
VI	L2	VE	(2)/ TAMENI/(1)/RU/
VI	N4	VE	(2)/ NI/(1)/RU/
VI	PB	VE	(2)(1)/RU/
VL	N4	VE	(2)/ DE/(1)/RU/
VL	RS	VE	(2)/ KOTODE/(1)/RU/
VT	AN	VE	(2)/ NI/(1)/RU/
VT	L1	VE	(2)/ E/(1)/RU/
VT	N3	VE	(2)/ O/(1)/RU/
VT	PA	VE	(2)/ NI/(1)/RU/
VT	RA	VE	(2)/KU/(1)/RU/
WN	PE	B2	(2)/RARETA// TOKI/

PATTERN (1)	(2)	SUBSTITUTE	CORRESPONDING-JAPANESE
PT	L1	PE	(2)/ E/(1)
PT	N3	PE	(2)/ O/(1)
PT	PA	PE	(2)/ NI/(1)
R1	PE	RS	(2)/IMASITA/
R1	PT	RS	(2)/IMASITA/
R1	VI	RS	(2)/RU/
RA	PA	AI	(2)/ NI/(1)
**	**	**	(1)(2)
TH	N4	N4	(1)(2)
TO	VF	L2	(2)/RU/
TO	VI	L2	(2)/RU/
UU	VE	SS	(2)
VD	AO	VE	(2)/ NI/(1)/RU/
VD	L2	VE	(2)/ KOTO O/(1)/RU/
VD	N4	VE	(2)/ O/(1)/RU/
VD	PB	VE	(2)(1)/RU/
VE	L2	VE	(2)/ TAMENI/(1)
VE	PB	VE	(2)(1)
VH	N3	VE	(2)/ O/(1)/RU/
VH	PA	VE	(2)/ NI/(1)/RU/
VI	AI	VE	(2)/KU/(1)/RU/
VI	AO	VE	(2)/ NI/(1)/RU/
VI	L1	VE	(2)/ E/(1)/RU/
VI	N3	VE	(2)/ NI/(1)/RU/
VI	PA	VE	(2)/ NI/(1)/RU/
VL	GE	VE	(2)/ITUTU/(1)/RU/
VL	PA	VE	(2)/ DE/(1)/RU/
VT	AI	VE	(2)/KU/(1)/RU/
VT	AO	VE	(2)/ O/(1)/RU/
VT	L2	VE	(2)/ KOTO O/(1)/RU/
VT	N4	VE	(2)/ O/(1)/RU/
VT	PB	VE	(2)(1)/RU/
WN	GE	B2	(2)/RU// TOKI/
XX	VE	SS	(1)(2)/ BESI/

Chapter 5

ALGORITHM OF MECHANICAL TRANSLATION

5.1 Algorithm of English-Japanese translation

The program of mechanical translation from English into Japanese consists of seven parts as shown in a block diagram in Fig.5.1.1. The following lines give the detail explanation of each program, and the next section will give a concrete example.

5.1.1 Reading of the original text.

Input sentences are supposed to be English, and constructed by 26 alphabets in capital letter and numerals and several symbols such as comma, brackets, hyphen, apostrophe, question mark, period, space, and a few symbols used in mathematics. When original sentences are punched on paper tape, some slight modifications are made. That is, a period which is attached to indicate abbreviation is ignored, and abbreviated words which are confusable with normal words are deformed. For example, AM (ante meridiem) is confused with am (linking verb), US(united state): us(pronoun), or MISS(mistress): miss (to make mistake) etc. In such cases

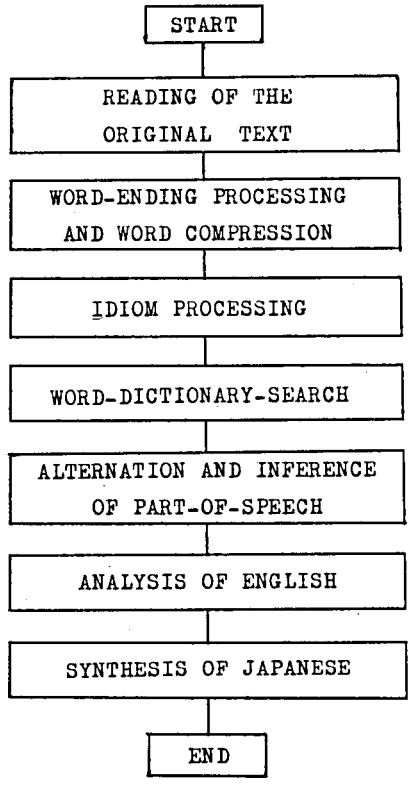


Fig. 5.1.1 Block diagram of mechanical translation.

a special symbol * is attached to the abbreviation words like AM*, US*, MISS*. Further, somewhat exceptional expressions such as "out-of-date" and "the present paper deals only with and and not type searches", are transformed into "out of date" or "...with and* and not* type...". In some cases, the pre-edition which is different from the above mentioned one is effective. The pre-edition, however, does not mean a complex one, but it treats a simple one such as the insertion of comma between very large phrases, or the deletion of redundant comma, or the substitution of syntactically ambiguous words with other equivalent words, for example, the underlined word in "...about five years..." is changed to "...nearly five years...". Such pre-edition can be done without stopping the text-punching. More detailed explanation about pre-edition will be given in section 6.4.

When a input sentence is fed into a computer from the paper tape reader(PTR), some special symbols which are not on the tape are inserted by the program. They indicate the beginning of a sentence (#Δ) and the ending of a sentence (Δ#) and a boundary between sentences (**).

(Original form) I AM A BOY . SHE IS A GIRL .

(Internal form) ** #Δ I AM A BOY Δ# ** #Δ SHE IS A GIRL Δ#

↑ beginning mark ↑ boundary mark ↑ ending mark

These symbols are useful in characterizing the functional role of the words situated near them, or preventing the influence of one sentence from going beyond the other sentence when many sentences are translated together, because all sentences which are read into a computer at a time are processed as if they were a single sentence. The number of words which can be treated at a time are about 90 words if syntactic

analysis trees need not be printed, and in the case of printing the tree, it must be less than 60 words. This limitation comes only from the number of core memories used in the experiment.

5.1.2 Word-ending processing and word compression.

When a sentence is being read into a computer from the paper tape, the end of a word or segment is recognized by space or punctuation marks including brachets, hyphen and apostlophe. Each mark is regarded as a single word, but space is not regarded as a word. For example, the next sentence which is punched in the same manner with ordinary type-writing is segmented into 23 words including boundary symbols.

MR BROWN'S HAT, WHICH IS OLD-FASHIONED, IS SELD AT 1 DOLLAR (360 YEN).

/ ** / # /MR/BROWN/ ' / S /HAT/ , /WHICH/IS/OLD/ - /FASHIONED/
/ , /IS/SELD/AT/ 1 /DOLLAR/ (/ 360 /YEN/) / # /

If a special sign indicating the end of an input sentence is read, then word-end processing mentioned in section 4.1 is applied to each word. That is, if a word has one of the endings shown in Table 4.1.1 in section 4.1, the longest one is taken off unconditionally. For example, "cities" has three endings "s", "es", and "ies", but the longest one "ies" is taken off from the "cities"; then it becomes "cit". The taken-off endings are reserved till the word dictionary is consulted, but in the case of four-letter-ending only the last three letters are regarded as word-ending. The remaining part of the word after the end-processing is applied is called a pseudo-stem, and segmented every five letters, then the segments are added each other simply as binary

number. As for a segment which consists of less than five letters, zeroes are inserted in the least significant figure. Some examples are shown below.

Original-form	End-processing	Segmentation	Compression
BOY	BO	BOOOO	BOOOO
EXAMPLE	EXAMPL	EXAMP/LOOOO	YXAMP
S	O	OOOOO	OOOOO
SYSTEMATICALLY	SYSTEMATICALL	SYSTE/MATIC/ALLOO	Q≠9:Q

Fig. 5.1.2.1 Word-end processing and compression.

Internal memories used for one word are 11 characters each as shown below,

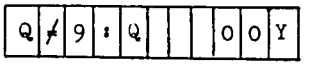
SYSTEMATICALLY ; 

Fig. 5.1.2.2 Internal form of an input word.

Five characters are used for the compressed stem, and three characters are used for the word endings, and if the word has no ending, this part is blank. The last letter of the word ending, if any, is regarded as a morpheme sign which represents the character of the word endings. The remaining three characters are not used in this step. The memory area is segmented every 11 characters and the information for the input words are stored into the "sub-area" in the same order as the input words.

The original words must be reserved to be used as the translation words when the input English words are not found in the dictionary.

5.1.3 Idiom processing.

It is supposed that there is such an idiom dictionary as mentioned in section 4.3, and each word in the read-in sentence is numbered 1,2, ... from the top of the sentence. A compressed pseudo-stem for the input words will be called "element", and a compressed form in the idiom and word dictionary will be called "head" or "key" or "entry", in the following discussion.

The main procedure in the idiom processing is to substitute a sequence of some elements with one element if the sequence is the same one stored in the idiom dictionary. Now, suppose that the ith element coincides with the kth head in the main word column of the idiom dictionary. Then investigate whether the 4th head in the kth row is blank or not. If it is not blank, then compare it with the i+1th element. If they coincide and the 2nd head in the kth row is not blank, the 2nd head is compared with the i-1th element. If they coincide and the 1st head is not blank, then the i-2th element is compared with it. In this manner the comparison is carried out. If a blank head appeares

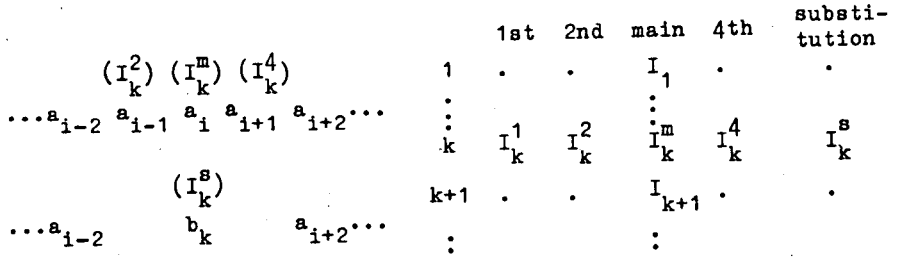


Fig. 5.1.3.1 Schematic expression of idiom processing.

in the course of comparison, the further comparison is unnecessary, and it is regarded as a successful comparison. If their comparisons are all successful, the sequence of elements which has the same struc-

ture as the kth idiom is substituted by the substitutional word in the kth row. For example; in Fig.5.1.3.1 above, if $I_k^1=\Delta$, $I_k^2=a_{i-1}$, $I_k^m=a_i$, $I_k^4=a_{i+1}$, $I_k^5=b_k$ the original sequence of element becomes "a_{i-2} b_k $a_{i+2}\cdots$." When an idiom is found and substitution takes place, the mark is transmitted to b as an ending mark of it, if the ending mark of a_{i-1} or a_i or a_{i+1} is G or D. This is to reduce the memory by eliminating a similar idiom which is different from the others only in conjugation, for example, "acount for" and "accounted for".

When the comparison with the kth row is unsuccessful, the ith element is compared with the k+1th head in the main word column. If this comparison is successful, regarding k+1 as k, the same procedures mentioned above are repeated, till there appears a head in the main word column which does not coincide with the ith element. This is because there are several idioms which have the same main words. If there are no idioms which have as main words, the i+1th element is compared with the main word column and the same procedure are performed.

By the idiom processing procedure, the sequence of elements is substituted with one element, and such information as part of speech and translation words are not given in this step.

5.1.4 Word-dictionary-search.

After the idiom processing, the word dictionary is consulted in order to get the information of part of speech and translations. The word dictionary for the mechanical processing must be divided into several pages because of its large amount of information, though it

depends on the memory size of the computer. Moreover the method of arranging head word must be taken into consideration with the relation to page division and search algorithm. To arrange words in alphabetical order as in an ordinal dictionary is not effective when such word compression is done. In this paper, all compressed words are arranged in ascending order regarding them simply as binary numbers of 5 x 6 bits, and this ordered entries are divided into several pages, to which the largest head word in the page is given as an index.

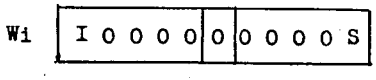
The binary numbers of the compressed word in the input sentence do not appear necessarily in ascending order; therefore they must be rearranged in ascending order not to diminish the search efficiency. The reason is that a word dictionary contains a very large amount of information, and they are stored in auxiliary memory such as a magnetic tape or magnetic disc or magnetic drum; therefore the number of open pages must be as less as possible in order to lessen the transfer time. However, if each element is searched in the order which it appears, it takes much time to search all elements.

To lessen the search time, it is necessary to investigate in advance to what pages each word belongs by the help of index, and then necessary pages must be opened in turn. If a certain page is opened, all the words which may be contained in the page are searched at a stretch independent of the input order. In this case, however, the relation between search time and necessary core memories must be considered. It requires many memories to list all the words which belong to the same page for each page. On the other hand, if each word is given the page-number to which the word belongs, less memories may be needed, but much search time will be necessary. Therefore as an intermediate method, the next method is adopted in this paper. First, the

head address of storage area for each word is regarded as the index for the word, and these indexes are put into the stack in the input word order. Second, the index of the first page of the word dictionary is compared with the elements by the aid of index in the stack. If there are no elements which belong to the first page, this page need not be opened. In this manner, the index of each page is compared with the elements in the sentence, and if there appears an element which may belong to a certain page, the page is opened, that is, the information contained in the page is transfered to the core memory. Then the element is searched in the page, and the index of the element is rejected from the stack. Once the page is opened, all the elements which have the possibility of belonging to the page are searched in the page independent of the input word order, and their indexes are all rejected from the stack. Therefore, after a certain page is closed, the remaining indexes in the stack represent the elements which are not yet processed, so the procedure to judge whether the element is searched for in the dictionary can be omitted. The same stack can be commonly used as the stack for the unprocessed elements; therefore the number of memories for the stack can be the same as the number of the input words, and the number is smaller than when each page has its stack.

In the following, the search algorithm for an open page will be explained. The head words in the dictionary are sorted in ascending order, so that it is only necessary to look for the coincident entries by the simple look-up method. But as a result of the word-end processing, some words have the same elements. In this case the entries of the dictionary have special forms as shown in section 4.4; therefore the taken-off word endings must be taken into consideration. Then the treatment of such a case will be explained in detail.

Now, suppose that an element (Wi) in the input sentence is compared with the dictionary entries in a certain open page by the simple look-up method, and then the kth head coincides with Wi. If the k+1th head is not a suffix entry, then the information of the kth head is for this word (wi). Here a suffix entry is judged by investigating whether the upper two figures of the entry are zero or not. A suffix entry always has ØØ*** form. If the k+1th word is a suffix, the suffix (Xi) of the element (Wi) is compared with the k+1th head. If they coincide with each other, the k+1th information is given to Wi. If Xi is larger than H_{k+1}, Xi is compared with H_{k+2}, and so on, till there appears such head (H_{k+j}) which coincides with Xi, or larger than Xi. If $Xi < H_{k+j}$, there are no heads, which coincide with Wi, in the following entries after H_{k+1}. Therefore Wi is regarded as equal to the kth head (H_k). If $Xi = H_{k+j}$, the information of H_{k+j} is given to Wi.



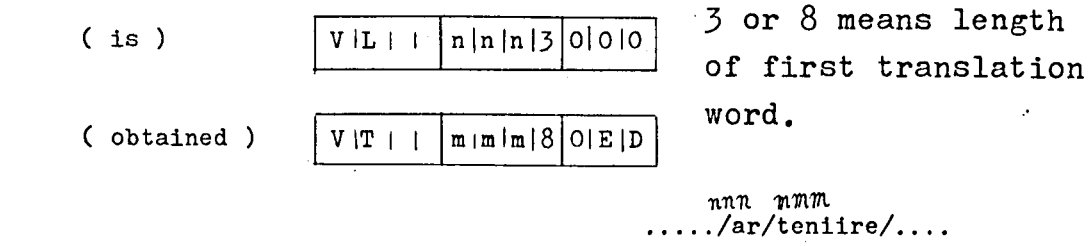
	head(H)	
k-1		
k	I0000	(N4;WATASI)
k+1	0000F	(C2;NARA)
k+2	0000S	(VL;AR)
k+3	IN000	(P1;NO NAKA)

Fig. 5.1.4.1 Dictionary search for the same head word.

In this method, however, the head which is smaller than 1 is confusable with a sufix entry, then for such words a special algorithm is considered. That is, if a suffix entry disturbs the normal ascending order, there must be at least one place where a small number comes after a larger one; therefore by investigating this point the boundary between a suffix and an ordinal head can be recognized.

If the element has the possibility of belonging to page i, and a coincident word can not be found in the page, the word is not registered in the dictionary. Most of such words are "proper nouns" or "technical terms", then their parts of speech are thought to be "noun"(N1), and the original English words themselves are given as their translations.

When a coincident entry is found, the information which the entry contains is given to the element. They are parts of speech and translations. Thus, after the word dictionary search the information contained in each sub-area is changed as follows,



As for the word ending, if it is referenced to find its entry, it is ignored, because the information of word ending is reflected already on the part of speech in such a case.

5.1.5 Alternation and inference of part of speech.

A part of speech which is obtained from the word dictionary is, as it were, a static one, and it may be different from the practical role in the sentence. Here the word "part of speech" only concerns the syntactic roles, and not the semantic function. For example, in "a rolling stone", the word "rolling" may be adjectival as a semantic

function, but it can be only said to be a ing-form of verb from the syntactic point of view. It is very important to distinguish the ing-form and past-form of verb to know the function of the verb. Then according to the word ending, the static parts of speech gotten from the dictionary are changed into rather dynamic ones. That is, such ending marks as D(ed, ied, ved), G(ing), and R(T)(er, est) have the possibility of changing the static part of speech, but the other endings do not affect the syntactic role. There are three transformation tables corresponding to the end marks D, G, R (Table 5.1.5.1).

Table 5.1.5.1 Transformation table for parts of speech according to the word-end-mark D,G,R.

Word-end-mark	Initial part-of-speech	Changed part-of-speech
D	(VX ,)	(PX ,)
	(N1,, VX)	(PX , N1)
	(AY , VX)	(PX , AY)
	(VX , VX)	(PX , PX)
	(?? ,)	(PX ,)
G	(VX ,)	(GX ,)
	(N1 , VX)	(GX , N1)
	(AY , VX)	(GX , AY)
	(VX , VX)	(GX , GX)
	(?? ,)	(GX ,)
R	(AI ,)	(RA ,)
	(AI , N1)	(RA , N1)
	(N1 , AI)	(RA , N1)
	(VI , AI)	(RA , VX)

where X = D, F, H, I, L, T, and Y = I, N, O.

If a word has one of the end marks D, G, or R(T), the part of speech of that word is searched in the transformation table corresponding to the end mark, and they are substituted by the transformed one. In the

table a priority between two parts of speech as well as morphological transformation is taken into consideration. For example, if a word which is given two parts of speech N1 and VT in this order has an end mark G, then G table is looked up for N1 VT, and they are substituted GT N1. Because the word ending "ing" indicates that the word is possibly used as verb of ing-form rather than as a noun. But if a word is given only N1, it is not changed because noun can not be thought to conjugate.

The above mentioned transformation is applied to the parts of speech which were obtained from the dictionary. But as for the word which is given temporally a "noun" symbol because it is not registered in the dictionary, its adequancy must be tested by another method. If a word which is found in the dictionary has "N1", it is not necessary to exchange its symbol even if its end mark is G or D. As for a word which is not found in the dictionary, however, it is probably adequate to adopt GT or PT or B1(adverb) rather than N1 according to their ending mark "ing" or "ed" or "ly".

This morphological inference of part of speech gives fairly correct results when the word dictionary contains a large number of words. If the dictionary is a small size and the word "bed" is not registered in it, its part of speech may be infered as a past form (PT) by morphological information. Therefore such inference is not necessarily reliable though helpful in some cases.

Then another inference which uses contexual informations is tried. Now, as a simple example, suppose the word "bed" is regarded as the past form of a verb because it was not registered in the dictionary. If the word appears in the circumstance "the bed is ", it is inferenced to be a "noun" because the word just behind a determiner is often a

noun. But sometimes a past form appears in the same circumstance "the ——ed ——", and so this is not also reliable. But if a farther context is considered, the inference of part of speech may become more accurate, though it will be never perfect.

As for words which have only one part of speech, it is only necessary to change their symbols by the word-ending information, and not necessary to infere from the context. Because in the dictionary all possible uses of the word is registered as parts of speech, so the fact that they have only one symbol means that in any sentence the words work according to the symbol. Their functions are determined in the context by syntactic patterns.

The most important problem in this step is to select the adequate one out of the two symbols which a word has. For example, in the sentence "the rapid search requests...", the word "search" and "requests" have VT Nl and Nl VT respectively, but it is difficult to judge from the syntactical context of which part of speech must be selected in this sentence. In this case, however, the part of speech for "search" may be probably Nl rather than VT because the word is situated just behind an adjective. Then in the above example the sequence of parts of speech becomes DT — AN $\begin{matrix} \text{VT} \\ \text{Nl} \end{matrix}$ $\begin{matrix} \text{Nl} \\ \text{VT} \end{matrix}$ ... Which of them (Nl or VT) should be selected for "requests" depends on several factors which differ from case to case. One of the methods is to use probability concerning symbol connection, that is, the probability of DT AN Nl Nl and DT AN Nl VT. According to the probability table which was obtained from several thousand words, the probability ratio of DT AN Nl Nl and DT AN Nl VT is 5 : 1. Then, from the statistical point of view, the part of speech for "request" is determined as a "noun". In the above example this result is correct,

but it is not always so. If there is no effective information to select one of two parts of speech, the first part of speech is to be used.

Another method to treat multi-part of speech is to make all possible combinations of parts of speech of the input words, and to analyze each one. For example, in the case of "DT A1 VT DT N1 ", there are
N1 N1 VT
eight possible combinations. If the analysis of a certain possible sequence is blocked, that combination is rejected. By this method, several plausible combinations can be obtained. But this gives no complete solution to the selection of part of speech at all.

In general if the wrong part of speech is given to the word, the syntax analysis does not go well, so the correct determination of part of speech is one of the most important problems in the mechanical processing of natural languages. The problem, however, can not be solved only by morphological information or the restricted context in syntax, and therefore semantic information and some flexible scanning method must be introduced.

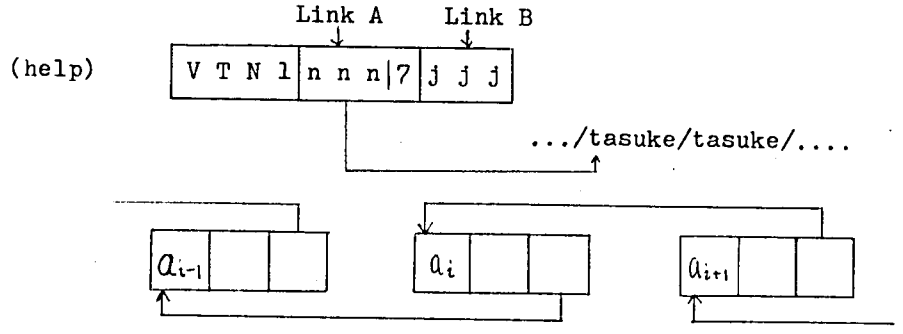
In this paper very simple inference rules are adopted which are shown in Table 5.1.5.2. These are simple context-sensitive rules whose form is ; $\omega(\alpha,\beta) \longrightarrow \omega(\beta,\alpha)$, if $\omega\cdot\beta$ is plausible, $\omega(\alpha,\beta) \longrightarrow \omega(\alpha,\beta)$, otherwise, where (α,β) is a set of part of speech given to a word, for example, "help": (VT,N1).

Table 5.1.5.2 Inference table for parts of speech

DP (VX , N1) ⟶ DP (N1 , VX) DP (VX , AY) ⟶ DP (AY , VX)
N1 (P4 ,) ⟶ N1 (P2 ,)
EA (N1 , VX) ⟶ EA (VX , N1) EA (AY , VX) ⟶ EA (VX , AY)

X= D, I, T Y= I, N, O DP= DT, P1, P2 EA= EE, VA

After this procedure, the information contained in the sub-area for each word becomes as follows



Each sub-area is connected by the link part (which will be called "connector"), and this list structure makes it easy to parse and make sub-translations. The connection order is reverse to that of input words, that is, the top of the list structure is the last word in the input sentence. This is because of the convenience for syntax analysis.

5.1.6 Syntax analysis

The main procedures in mechanical translation are syntactic analysis of source language and synthesis of target language, and these procedures are carried out by the use of rewriting rules (patterns) shown in Table 4.5.1 in section 4.5. Patterns have context-free forms, and its application algorithm is very simple. But they are classified into several hierarchies, therefore the application of them takes some steps, which are explained below.

The string of part-of-speech of the given sentence is supposed to be as follows, where a_i is a part of speech for ith element.

$$a_1 \quad a_2 \quad a_3 \quad \cdots\cdots \quad a_i \quad \cdots\cdots \quad a_{n-2} \quad a_{n-1} \quad a_n$$

First step: In the first step only A-3 and A-2 patterns are used. At first, three symbols are taken out from the end of string, that is, $a_{n-2} \cdot a_{n-1} \cdot a_n$ is taken out. Then A-3 table is simply looked up to search whether there is a pattern which has $a_{n-2} \cdot a_{n-1} \cdot a_n$ as a head part. In each table the patterns are in the same hierarchy, so they are sorted with the head part in ascending order. If there is no coincident pattern in Table A-3 , $a_{n-1}\ a_n$ is taken and searched in Tabel A-2 . The patterns which are searched in Table A-2 are the sub-strings of 3-symbol patterns, as shown below, so the patterns in A-2 can be said practically to be context-sensitive, though their form is that of context-free patterns. If $a_{n-1}\ a_n$ is not found in

$$a_{n-2} \quad \underline{a_{n-1} \quad a_n} \longrightarrow a_{n-2} \quad \underline{b}$$

A-2 Table, the next three symbols $a_{n-3}\ a_{n-2}\ a_{n-1}$ are taken out and searched in A-3 Table. If they are not in A-3 , $a_{n-2}\ a_{n-1}$ are

$$\cdots\cdots \quad a_{n-5} \quad a_{n-4} \quad a_{n-3} \quad a_{n-2} \quad a_{n-1} \quad a_n$$

Fig. 5.1.6.1 Parsing order.

searched in A-2 . The same process is continued in the manner shown in Fig.5.1.6.1. If in the course of table-looking-up there is a pattern whose head part coincides with a sub-string $a_{i-1} \cdot a_i \cdot a_{i+1}$ in the original string, $a_{i-1} \cdot a_i \cdot a_{i+1}$ is substituted by the substitution symbol indicated by the pattern, and a_{i-1}, a_i, a_{i+1} are arranged according to the word order indicator, and indicated particles are inserted. This procedure is shown in Fig.5.1.6.2.

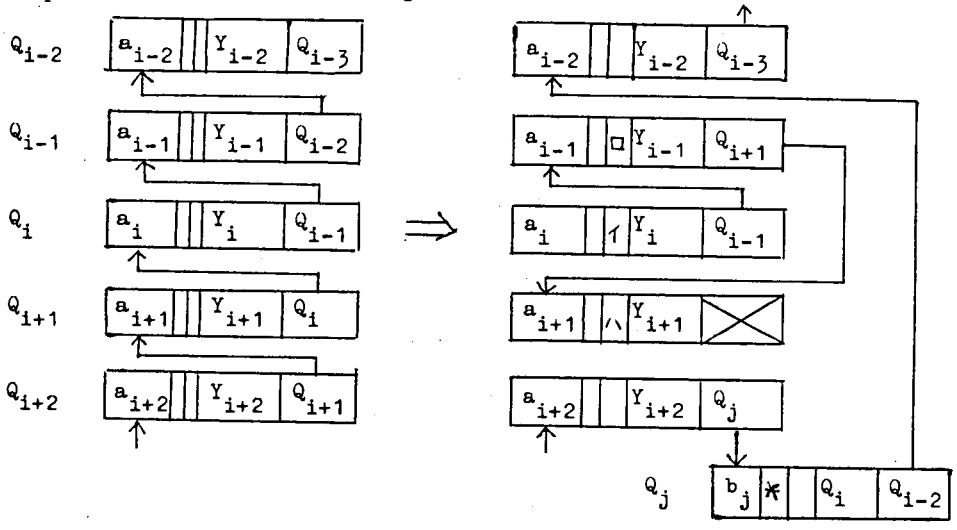


Fig. 5.1.6.2 Reformation of linkage by application of $a_{i-1} \cdot a_i \cdot a_{i+1} \longrightarrow b_j$ (2 1 3 ,1·□·/\)

That is, first, the links between a_{i-2} and a_{i-1}, a_i and a_{i+1}, a_{i+1} and a_{i+2} are cut off, and the new symbol b_j is inserted between a_{i-2} and a_{i+2}. Second, a_{i-1}, a_i, and a_{i+1} are linked according to the word-order indicator, the link part of the last word being make blank. Third, the particles are inserted in the indicated order. To distinguish the terminal symbol and the non-terminal symbol, 1 bit information(*) is added to the newly inserted container for the non-terminal

symbol. The sub-area (or container) for the newly inserted symbol(b_j) has two link parts. One is an ordinary link (connector) which makes the main string of the sentence, and the other link indicates the top of the connected symbol, the latter link indicates the top address of its translation words.

If the above mentioned procedure is continued till there appear no patterns, in A-3 and A-2 , which have the same head as the sub-string of the sentence, it goes into the second step. But if the second step was already passed, it goes into the third step.

Second step: In the second step, some parts of speech in the main string are changed. That is, N1 is changed to N4, and VH and VF are changed to VT, if interrogative snetences are not considered. And again it goes to the first step. The reason is as follows. If N1•PA $\longrightarrow$ N4 exists in A-2 , "DT N1 PA" become " DT(N1 PA) ". But this is not correct. It must be (DT N1) PA . On the contrary, if that pattern is rejected from A-2 , "N1 PA" can not be parsed. In the latter case, however, if "N1 PA" is changed to "N4 PA", it will be correctly parsed using A-2 patterns. Therefore such a change of part of speech is applied instead of storing rather irregular case patterns. This procedure is equivalent to dividing A-3 and A-2 into sub-classes. As for A-2 class, for example, the sub-class A-2_1 includes N4•PA$\rightarrow$N4(2•1, N1(NO)•ϕ), and the sub-class A-2_2 includes N1•PA$\rightarrow$N4 (2 1, N1(NO)•ϕ), and A-2_1 is in a higher hierarchy than A-2_2 .

It is possible to infer and alternate parts of speech by looking somewhat wider context in this step. For this purpose the connection table of parts of speech which is given in Appendix B will be useful.

Third step: In the third step, the pattern table B-3 and B-2 which are in a lower hierarchy than A-3 and A-2 are used. The procedure is almost the same as that of the first step except some process after substitution. That is, three symbols are taken out from the end of string, and searched in B-3 table. If it is not in B-3 , two symbols are taken out and searched in B-2 , and so on. If a coincident pattern is found, the quite same substitution, however, the B tables are not looked up continuously, but A table is used for the string which includes newly substituted symbol, as in Fig.5.1.6.4.

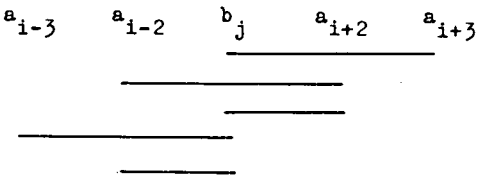


Fig. 5.1.6.4 Pattern matching after substitution.

If there is no coincident pattern in A Tables, the look-up procedure returns to table B again, and the same procedure is continued.

If the length of string for a given sentence becomes only one, it is said that the original sentence is correctly translated structually. But this does not necessarily mean that the translation is also correct from the semantic point of view. Only it can be said that if each word is substituted with an appropriate word, the translation may become a semantically correct one.

If the length of string does not become one after the application of B and A patterns, N4 in the remaining string is changed to N5,

and the third step is applied again. This is equivalent to the subdivision of B patterns.

Notwithstanding the change of part of speech and the reapplication of the third step, if the length of string does not become one, it is said that the sentence can not be analyzed or translated by this system. Even in this case, however, corresponding Japanese sentence is obtained, which is perhaps wrong as a whole but partially correct. The block diagram of syntactic analysis and synthesis is shown in Fig.6.1.6.5.

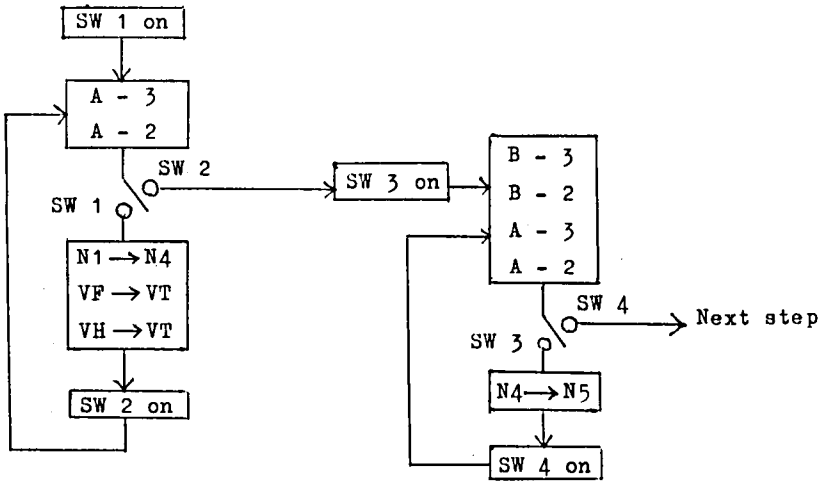


Fig. 6.1.6.5 Block diagram of pattern processing.

5.1.7 Synthesis of Japanese

It is very easy and simple to synthesize Japanese, because a partial translation has already been synthesized at the stage of pattern processing, where the word order is changed and the particles are inserted and each word is connected by the links in Japanese order. Then it is only necessary to print out in turn following the links. For this purpose, a push down store is used to simplify the algorithm. In this section, the sub-area which consists of 11 characters and con-

tains some information about the terminal or node points in the tree structure of the sentence will be called "container", which is shown in Fig.5.1.7.1 by a rectangular box.

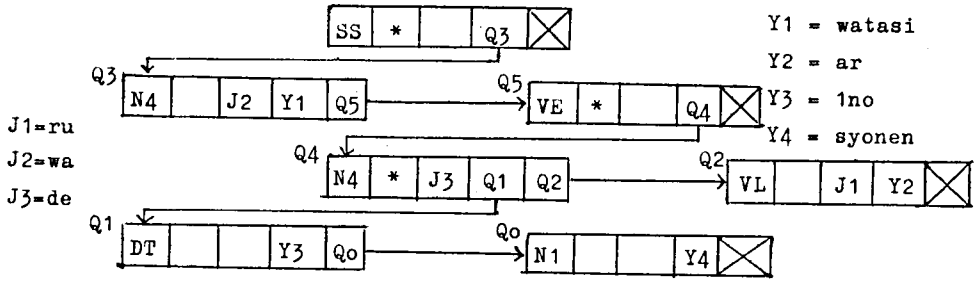


Fig. 5.1.7.1 List structure of synthesized Japanese.

At first, the last container which is make at the final step of pattern processing is put into the push-down store(Fig.5.1.7.2-a). Then it is investigated whether it is a terminal container or a non-terminal container. If it is non-terminal, the container which is linked by the link A part is put into the push-down store and the link A is erased (Fig.5.1.7.2-b).

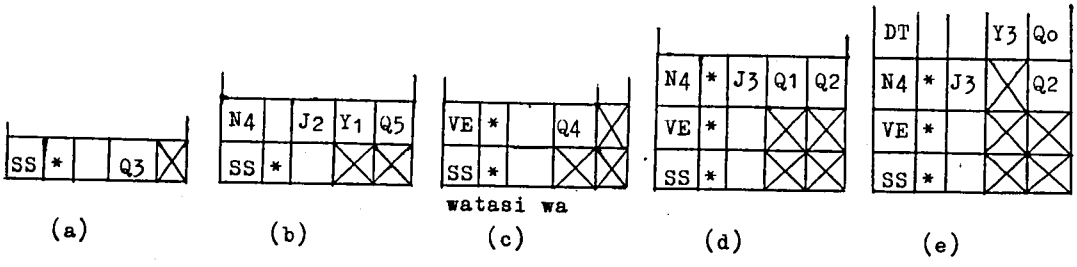
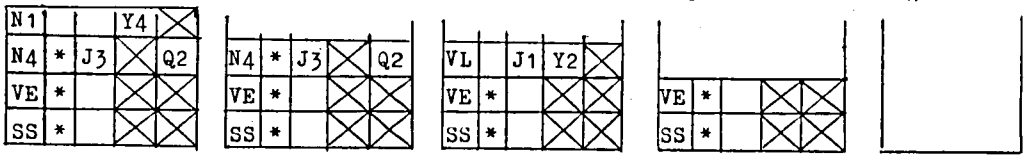


Fig. 5.1.7.2 State of push down store.

Then the same procedure is applied to the up most container in the push-down store. If the up most container is a terminal one, Japanese in Roman style is taken out from the address which is indicated by the link A part. After Japanese is taken out, the up most container is

investigated whether it has a particle symbol. If it has a particle symbol, the symbol is interpreted and the particle is taken out and placed immediately after the previous translation words. In this case if the particle is an inflection part of a verb or an adjective, its particle may be necessary to be changed slightly according to the last letter of previous words. It is explained later in this section about this change. After a particle is taken out, the link B part is investigated whether it links another container or not. If it does not link an another container, that is, the link B part is blank, the all processing concerning to the up most container is finished and it is rejected from the push down store. If the link B part is not blank, the container linked by its part substitute the top container (Fig.5.1.7.2-c).

In short, always the top most container of the push down store is only the object of processing, and in a container, the terminal or non-terminal indicator is first investigated, and the link A part is second, and then a particle symbol, and next the link B part is investigated.



(f) (g) (h) (i) (j)

Fig. 5.1.7.2 (continued)

If it is non-terminal and the link A part is blank and a particle symbol exists, the corresponding particle is taken out. When the link B part is investigated, if it is blank, the container is rejected, but if it is not blank, the container indicated by the link B substitute the top-most container. This procedure is continued till the push down store becomes vacant, and at this point the synthesis of Japanese

is completed.

Next, the rules of conjugation in Japanese will be explained. These rules can be regarded as a transformation rule.

As mentioned previously in section 4.4, the translation words of verbs and adjectives are stored in the dictionary only with its stem parts, and the information concerning conjugation types is not given to each word. Therefore the conjugation in Japanese must be carried out using only the information of the stem ending and particles to be connected. The inflection particles are distinguisehd from the other particles by their first letter. That is, the inflection particles which are shown below begin with R, I, A, or K, but the other particles begin with space mark or other letters. (ΔWA, ΔDE, ΔNI(NO) etc. Δ means space).

RU/	RARERU/	RARETA/	RARE/	RU BESI/	RU KOTO/
RU TO/	RU MONO/		IMASITA/	II/	ITUTU/
IOERU/	IOE/	I MONO/	I NI/	ITUTUARU/	
ANAI/	AN/	ASIMEYO/	ANAKATTADES/		KU/

General rules are as follows:

(1) If the last letter of the previous word is a vowel, the first letter of the inflection particle must be a consonant. If the last letter of the previous word is a consonant, the first letter of the inflection particle must be a vowel. The inflection particle which is indicated by the rewriting rule is one of the above particles. Therefore these forms must be changed according to the last letter of the previous word. But it is very simple. That is, if the first letter

of the inflection particles is not a desired one, the first letter is ignored. For example, if the indicated particle "RU" must begin with a vowel, the first letter "R" is ignored and it becomes "U".

(2) If the last letter of the previous word is "S", and the particle is "RU", then "U" is inserted between them. (-S + RU ⟶ -S + U + RU).

(3) If the last letters of the previous word are "SU", and the particle is "RU", the particle is ignored, and if the particle is not "RU", the last letter ("U") of the previous word is ignored.

(4) If the last letters of the previous word are "KU", and the particle begins with "A", the last letter "U" is changed to "O". And then rule (1) is applied. If the particle begins with "I", the last letter "U" is ignored.

(5) If the first letter of the particle is "A", and the last letter of the previous word is "S", "A" is changed to "E" (-S + AN ⟶ -S + EN). Some examples are shown in Tabel 5.1.7.1

Table 5.1.7.1 Some examples of conjugation.

KAK + RU ⟶ KAK + _U ⟶ KAKU

KAK + AN + RU ⟶ KAKAN + _U ⟶ KAKANU

KAK + ITUTUAR + RU ⟶ KAKITUTUAR + _U ⟶ KAKITUTUARU

KAK + RARE + RU ⟶ KAK + _ARE + RU ⟶ KAKARE + RU ⟶ KAKARERU

KAK + IMASITA ⟶ KAKIMASITA

ATAE + RU ⟶ ATAERU

ATAE + AN + RU ⟶ ATAE + _N + RU ⟶ ATAEN + _U ⟶ ATAENU

ATAE +ITUTUAR + RU ⟶ ATAE + _TUTUAR + RU ⟶ ATAETUTUAR + _U ⟶ ATAETUTUARU

ATAE + RARE + RU ⟶ ATAERARE + _U ⟶ ATAERARERU

ATAE + IMASITA ⟶ ATAE + _MASITA ⟶ ATAEMASITA

Table 5.1.7.1 (continued)

KENKYUS + RU ⟶ KENKYUS + U + RU ⟶ KENKYUSURU

KENKYUS + AN + RU ⟶ KENKYUS + EN + RU ⟶ KENKYUSEN + _U
⟶ KENKYUSENU

KENKYUS + RARE + RU ⟶ KENKYUS + _ARE + RU ⟶ KENKYUSARE + RU
⟶ KENKYUSARERU

DAMASU + RU ⟶ DAMASU + __ ⟶ DAMASU

DAMASU + AN + RU ⟶ DAMAS_ + AN + RU ⟶ DAMASAN + _U ⟶ DAMASANU

KU + RU ⟶ KURU

KU + AN + RU ⟶ KO + AN + RU ⟶ KO + _N + RU ⟶ KON + _U ⟶ KONU

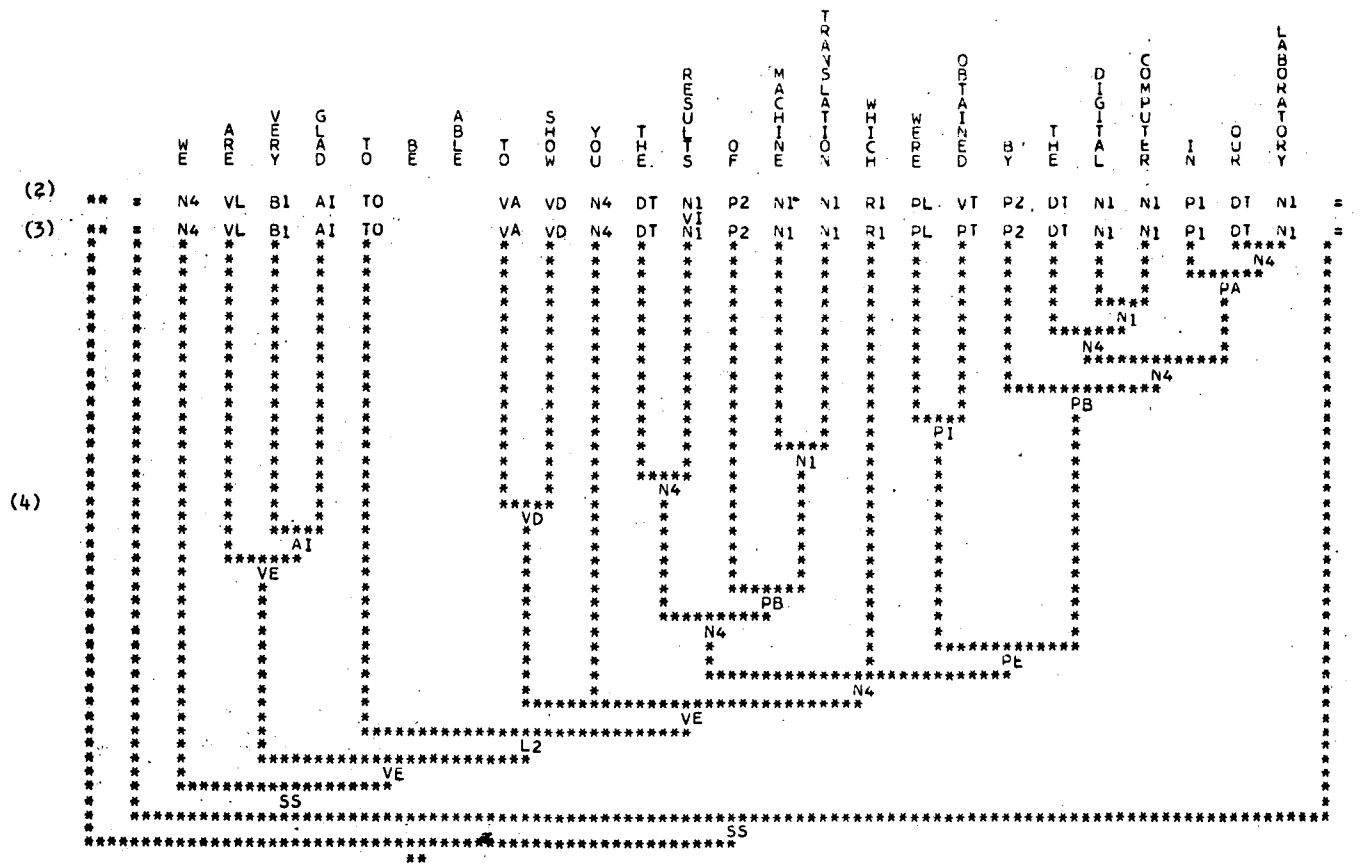
KU + ITUTUAR + RU ⟶ K_ + ITUTUAR RU ⟶ KITUTUAR + _U ⟶ KITUTUARU

5.2 Concrete Example.

A concrete explanation is given below using a sample (Fig.5.2.1) which was translated by the algorithm mentioned in the previous section, and printed out by the computer itself.

First, the input sentence which is shown in Fig.5.2.1-(1), is punched on paper tape and fed into the computer from PTR. The boundary mark (**) and the beginning mark (#△) are not punched on the input tape, but they are inserted by a program. Each word is separated into a pseudo-stem and a word ending, and each pseudo-stem is compressed to five characters. When a period is read in, it is replaced by the ending mark (△#). If the other sentences continue, the sentence boundary mark (**) and the beginning mark (#△) is inserted and the sentences are read in. Each word becomes as follows by the word ending processing.

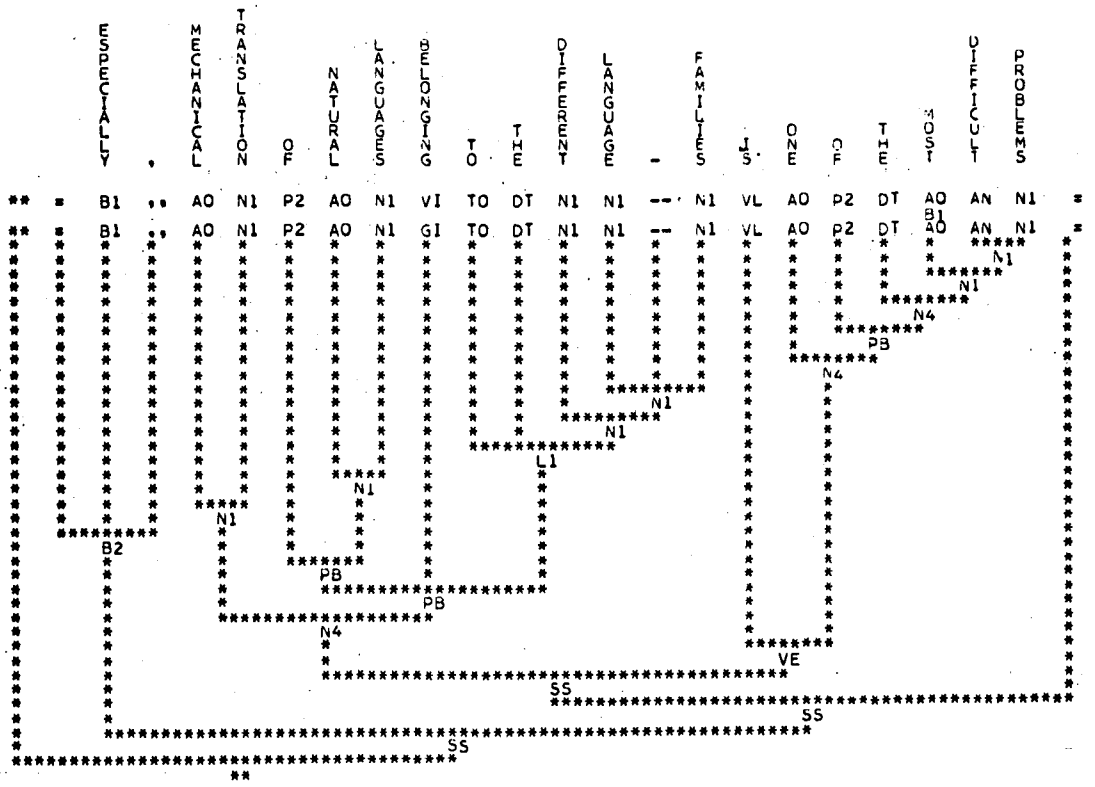
ENGLISH WE ARE VERY GLAD TO BE ABLE TO SHOW YOU THE RESUL-
(1) TS OF MACHINE TRANSLATION WHICH WERE OBTAINED BY THE D-
IGITAL COMPUTER IN OUR LABORATORY .

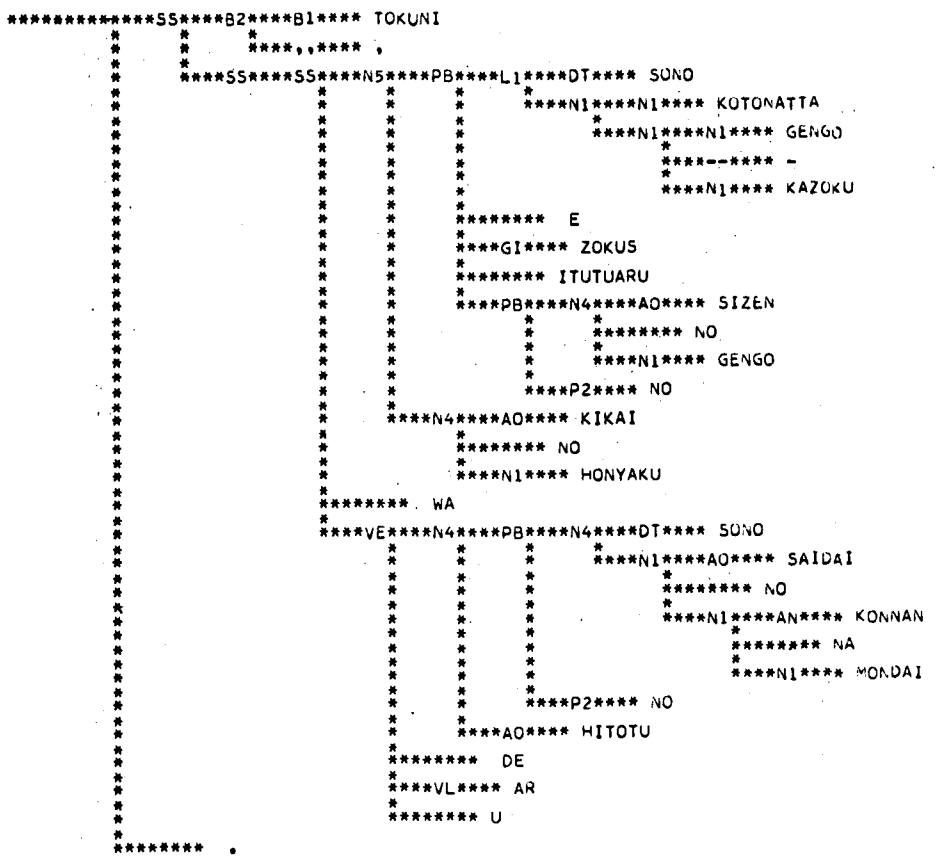


ニホンゴ ワレワレワアナタニワレワレノジッケンシツノナカニ(ノ)ソノジジタル
コンピュータデニイレラレマシタキカイホンヤクノソノケッカヲシメス コトガデ
キルタメニタイヘンウレシイ .

Fig. 5.2.1 An example for explanation.

ENGLISH ESPECIALLY, MECHANICAL TRANSLATION OF NATURAL LANGUAGES BELONGING TO THE DIFFERENT LANGUAGE-FAMILIES IS ONE OF THE MOST DIFFICULT PROBLEMS.





NIHONGO TOKUNI , SONO KOTONATTA GENGO - KAZOKU E ZOKUSITUTUARU SIZENNO GENGO NO KIKAINO HONYAKU WA SONO SAIDAINO KONNANNA MONDAI NO HITOTU DE ARU .

Fig. 5.2.2 A concrete example.

ニホンゴ トクニ , ソノ コトナッタ ゲンゴ - カゾク エ ソクシツツアル シゼンノ ゲンゴノ キカイ ホンヤク ワ ソノ サイダイノ コンナン モンダイ ノ ヒトツ デ アル .

W/E , AR/E , VER/Y , GLAD/ , TO/ , B/E , ABL/E , TO/ ,
SHOW/ , YOU/ , TH/E , RESULT/S , O/F , MACHIN/E , TRANSLATION/ ,
WHICH/ , WER/E , OBTAIN/ED , B/Y , TH/E , DIGITAL/ , COMPUT/ER ,
IN/ , OUR/ , LABORATOR/Y

The left side of the slant is a pseudo-stem, and the right side is a word-ending. The pseudo-stem is segmented per five characters, and each segment is added as a binary number. Their compressed word (element) becomes as follows.

WØØØØ	ARØØØ	VERØØ	GLADØ	TOØØØ
BØØØØ	ABLØØ	TOØØØ	SHOWØ	YOUØØ
THØØØ	)ESUL	OØØØØ	9ACHI	,,4$H
WHICH	WERØØ	#BTAN	BØØØØ	THØØØ
N(GIT	6OMPU	INØØØ	OURØØ	V4Z&R

Among the above words, there are the same words BØØØØ, one is for "BY", the other is for "BE". But in the word dictionary they are stored as Fig.5.2.2, and so the confusion can be avoided by use of word-ending.

Now, the idiom dictionary is searched. In this example, "be able to" is found in the dictionary and is substituted by "can" is not printed in Fig.5.2.1, but in the memory "be able to" is substituted by "can". There is only one idiomatic expression in this example.

BØØØØ	(for be)
ØØØØY	(for by)

Fig.5.5.2

Next, each word is looked up in the word dictionary, and its part of speech and translation words are taken out. In Fig.5.2.1-(2), the part of speech for each word is shown beneath each word. The part of speech of "we" is N4, "ARE" is VL,..., and "be able to", which is

equal to "can", is given VA. To the word "results", two parts of speech N1 and VI are given. The order of N1 and VI is the same as that of an ordinary English dictionary, that is, N1 is the first and VI is the second. But the actual role in the given sentence is determined by the word ending and syntactic context. The word "obtained" has "VT" because its pseudo-stem "obtain" is the same form as present tense form. After the word dictionary is consulted, the alternation and inference of part of speech are tried. It is only "VT" for "obtained" which must be changed by the word ending information, that is, VT becomes PT because of the ending mark D(ed). As for "results", its part of speech is determined as "N1", because the word just behind the determiner is probably N1(noun) rather than VT(present verb) in almost all cases. On the contrary if a word having N1·VI appears just behind an auxiliary verb, its part of speech will be regarded as VI rather than N1. In this example, there are no other words which need the changing of part of speech. As a result, the string of part of speech for the input sentence becomes as shown in line (3) in Fig.5.2.1.

Then, the syntactic analysis and synthesis of the source and target language are carried out. First, the three parts of speech DT, N1, and $\Delta\#$ are taken out from the end part of the main string, and searched in A-3 Table by simple table looking-up. There is no pattern which conincides with this. Then N1·$\Delta\#$ is searched in A-2 table. This is also not found. Then the next substring P1 DT N1 is searched in A-3 . Again this is not found. Then DT N1 is searched in A-2 . This is found in the table, that is, DT N1 $\longrightarrow$ N4 (1·2, ~~0·0~~). This rule indicated that DT N1 is substituted by N4, and the word order in Japanese is the same as that of English (1 2, ~~0·0~~), and no particles are inserted (1·2, ~~0·0~~). After the substitution the

parsing procedure is again continued in the same manner, but this time, the substring is taken out so that the newly substituted symbol is included in it. In the example, the sequence of substring are P1 · N4 · Δ#, N4 · Δ#, N1 · P1 · N4, and P1 · N4 . They are searched in A-3 , A-2 , A-3 , and A-2 in that order. The pattern P1 · N4 is found in A-2 , and it is substituted by PA (prepostitional phrase) according to the rule P1 · N4 ⟶ PA (2·1, Ø·Ø·Ø). Then the same procedure is applied till the first symbol of the main string comes to be included in the sub-string. If the length of string is not one at the end of the above procedure, the symbol N1s in the remaining string are changed to N4, and again the same procedure is applied to the remaining string. As for this example, there are no N1s in the remaining string (** #Δ N4 VE TO VD N4 R1 PI PB Δ#), and so the changing process has no effect.

The next step uses B-3 and B-2 table in place of A-3 , A-2 , and the parsing procedure is almost the same as the previous one. That is, the three part of speech PI, PB, Δ# are taken out from the end part of the remaining string, and searched in B-3 . This is not found in B-3 , and then PB·Δ# are searched in B-2 , and so on. The pattern "PI · PB" is found in B-2 , and processed by the rule PI PB ⟶ PE (2·1 , Ø·RU·Ø). After this processing, the pattern table A-3 and A-2 are used again for the sub-string including the newly substituted symbol PE. If there are coincident patterns in A table, that table is used in succession. If A table any more contains the coincident pattern with the sub-string, again B table is used. In this example, N4 · RI · PE is found in A-3 , and so A table is used in succession. The rule N4 · R1 · PE ⟶ N4 (3·1·0 , IMASITA·Ø·Ø) treats a relative clause which plays an adjectival role. It must be noted that complex

sentences including relative clauses can be analysed by these rules as if they were simple sentences.

Thus in the third step, A and B tables are used. Finally the whole strings become only one symbol "**". This time the translation into Japanese is already made, and it is only needed to print out using the push-down store.

The tree structure of Japanese will suggest how the push-down store is used. The symbols at the tree nodes and tip of branches are parts of speech of English words, and it can be clearly understood that what part of English corresponds to which part of Japanese from the comparison of the English tree and Japanese tree. In the Japanese tree, the terminal words connected to the branches which have no part of speech are particles which are inserted by patterns.

As for the conjugation of verbs and adjectives, several examples are shown in the Japanese tree. For example, the rule PL · PT ⟶ PI (2·Ø , RARE·Ø·Ø) indicates that "TENIIRE" is to be followed by "RARE". In this case, "RARE" is connected to "TENIIRE" because the last letter of "TENIIRE" is a vowel and the first letter of the particle is a consonant. On the other hand, the rule N4 · Rl · PE ⟶ N4 (3·1·Ø , IMASITA ·Ø·Ø) indicates "IMASITA" is connected to " RARE"(PE). But the last letter of the previous word (RARE) is a vowel, so the first letter of the particle which is a vowel is ignored. Then it becomes " RAREMASITA". As for the rule VL · AI ⟶ VE (2·Ø , II·Ø) , "II" must be connected to " URESI"(AI), but the last letter of the previous word (URESI) is a vowel, so the first letter of the particle which is a vowel is ignored, and then it becomes " URESII".

A translation (6) in Roman letters is composed by gathering the right most words of each branch in that order. A translation in KANA

letter is gotten from (6) by simple transformation of Roman letters into KANA. This transformation is not difficult, but there are some cases where the syllabic nasal (N) is confusable with the contracted sound and the N series (NA, NI, NU, NE, NO). For example,

H O N Y A K U　　T A N I
ホ ン ヤ ク　　タ ン イ

H O N Y A K U　　T A N I
ホ ニャ ク　　タ ニ

HO/ NYA/ KU and HO/ N/ YA/ KU can not be distinguished from the morphological point of view. Therefore it must be written as HONIYAKU if it needs to be read HO/ NYA/ KU. As for TANI, it must be written TANWI, if it needs to be transformed タンイ .

Chapter 6

EXPERIMENT

6.1 Mechanical translation system

An experiment has been carried out using a rather small size general purpose digital computer, NEAC 2200/200 (a family of Honeywell computer). The specification of this computer is shown in Table 6.1.1.

Table 6.1.1 Specification of the computer used in the experiment

Central processor	NEAC 2200/200	Peripherals	
Minimum data unit	6 bits (1 ch.)	Magnetic tape	20,000 ch/sec
Instruction	Variable length	N204B1	556 bit/in
	Two address	(5 units)	
Main memory (core)	16,384 ch.	Paper tape reader	300 ch/sec
Index register	6	N209A1	
Memory cycle time	2 μs	(1 unit)	
Kind of instructions	38	Paper tape puncher	60 ch/sec
Arithmetic	8	N210A1	
Logic/Condition	8	(1 unit)	
Control	15	Line printer	420 line/min
Input/Output	2	N206A1	120 ch/line
Interrupt	4	(1 unit)	(alphabets,kana numerals, signs)
Edit	1		
Processing speed	Comparison 38 μs		
(5 ch.)	Addition 48 μs		

The main core memory is only 16 Kch, and it can not contain all programs and dictionaries, so magnetic tapes are used as auxiliary memories. The composition of equipments is shown below.

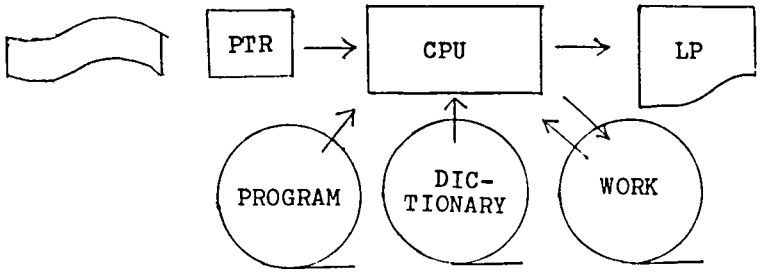


Fig.6.1.1.

The program is written in assembler language EASYCODER, and consists of about 1200 statements (about 7200 ch), and is divided into five segments because of the limitation of memory size. The size of each segment and various dictionaries are shown in Fig.6.1.2.

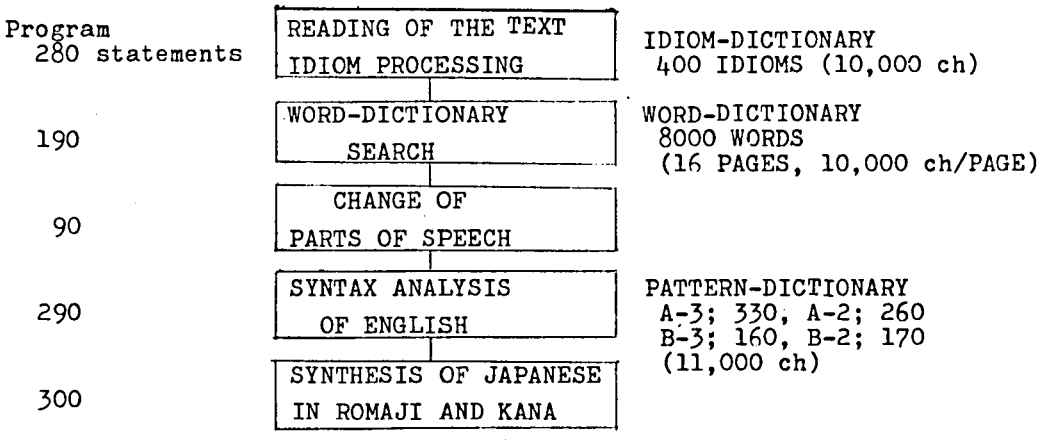


Fig.6.1.2 The size of each segment and various dictionaries

The translation time is about 0.5 - 1 second per word, and its distribution is shown below (Fig.6.1.3 shows a time table for a sentence composed of 45 words). If there are many core memories, say about 250 Kch, then the translation time will become about 0.3 - 0.6 second per word, or 5000 - 12000 words per hour. At this rate, one page of technical paper which includes 1000 words can be translated in five or twelve

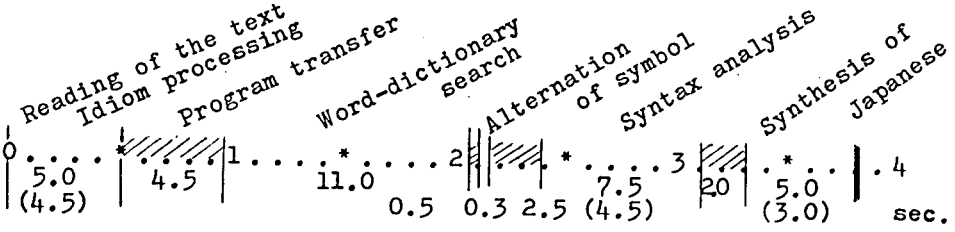


Fig. 6.1.3 Time table for translation of 45 words-sentence (bracketed time is for 30 words-sentence).

minutes. If the cycle time of the computer becomes much faster, the translation time will become one-tenth or less. Therefore as far as translation time is concerned, mechanical translation is far faster than human translation and will be of practical use.

The speed-up of translation time is achieved by the separation of program and grammatical information and by the adoption of the simple table look-up method. These techniques make it easy to study the effectiveness of several grammars by alternation of rewriting rules and parts of speech without affecting the program itself.

Natural language, however, is not so stiff that it seems to be rather unnatural to treat all parts of processing by fixed phrase structure like patterns. Especially in idiom processing, selection of part of speech, or treatment of adverbial words, their structures are different from case to case; therefore a more complicated procedure may be needed. But judgement of "when" and "which" in application of several kinds algorithm is very difficult for a machine; therefore it takes much time, if case by case treatments are performed. So human intervention is necessary in mechanical processing of natural languages at a certain stage. The purpose of the experiment described in this paper is to investigate to what extent such a hard and fast rule can be applied to natural languages.

6.2 Samples for translation

Sample sentences for mechanical translation are selected from several fields at random. This translation system makes scientific papers or their abstracts an object of syntactic study, paper in other fields being taken up as a reference. Such samples are as follows.

(1) Scientific papers

(a) B.G.Lamson and B.Dimsdale, "A natural language information retrieval system", Proc. of IEEE,vol.54,no.12,pp1636-1640, 1966.

(b) Abstracts of 19 papers appeared in Proc. of IEEE,vol.55, no.3, 1967.

(c) T.Toshihiro, "Germanium hyper abrupt varactor", -ABSTRACTS- Jour. Inst. Elec. Comm. Engrs. Japan, vol.49, no.2, pp[illegible]-3, Feb. 1966.

(d) H.Elliot,et al., "Satellite observations of the energetic particle flux produced by the high-altitude nuclear explosion of July 9, 1962", NATURE , no.4848, pp1245-1248, Sept. 29, 1962.

(2) News papers

(a) "UNSC calls for immediate mideast truce", see (b).

(b) "EDITORIAL", The mainichi daily news, June 8, 1967, page 1.

(c) "Taiho's Europe trip may hurt next performance", The Japan times, June 8, 1967, page 7.

(3) English text of middle school

W.L.Clark, "The junior Crown English course 1 ", SANSEIDO, 1967.

(4) Essay

A.G.Gardner, "A man and his watch", from 'Many furrows', 1924.

In order to clarify the syntactic difference between these sample sentences from different fields, the length of a sentence (the number of words in a sentence) and degree of complexity are calculated for each group of sentences, and they are compared. Complexity degree is defined as the number of words which cause the inversion of word order in Japanse or require the insertion of particles. Such words (complex

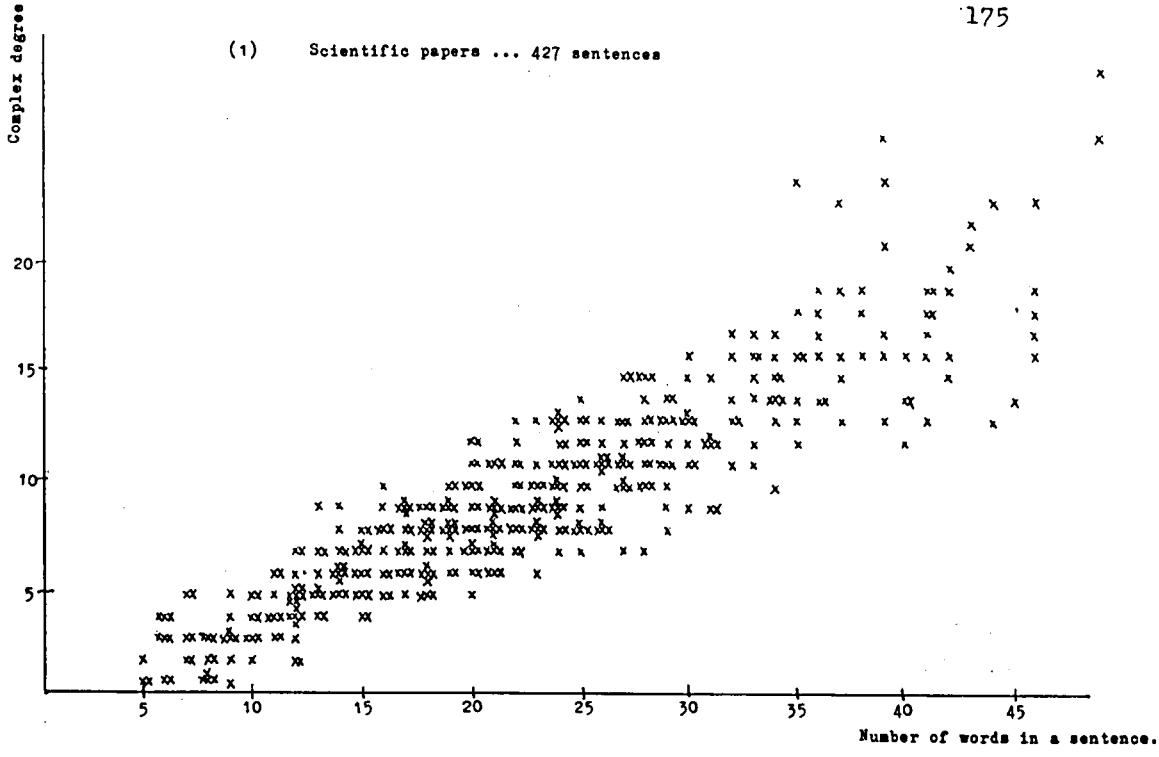
words) are mainly function words and verbs; these words whose parts of speech are DT, N1, N3, N4, AI, AN, AO, B1 are not regarded to contribute to the complexity degree.

Several examples of the calculation of complexity degree are shown in Fig.6.2.2. In example (1), only one word "am"(VL) belongs to the complex word class, so the complexity degree of sentence (1) is 1. As regards sentence (2), two words "have"(verb) and "in"(P1) belong to a

		Complex degree
(1)	I am a boy.	
	N4 VL DT N1	1
(2)	I have a book in my hand.	
	N4 VH DT N1 P1 DT N1	2
(3)	This is a result which was obtained by the computer.	
	N3 VL DT N1 R1 PL PT P2 DT N1	5
(4)	This is a result obtained by the computer.	
	N3 VL DT N1 PT P2 DT N1	3 + 2 = 5

complex word group, so the complexity degree of (2) is 2. The complexity degree of (3) is 5. But past form verbs or ing-form verbs are different in understanding from present forms when they are used as adjectival words which modify noun phrase from the right side. Then in that case, complexity degree of such word gets another two points. For example, sentence (4) gets only three points in ordinary counting, but the verb "obtained" works as an adjectival word, it gets another two points. The judgement of the role of verb is performed by man. The results of complexity calculation are shown in Fig.6.2.3.

This complexity factor reflects in a certain degree the real difficulty Japanese people have in understanding English. The complexity



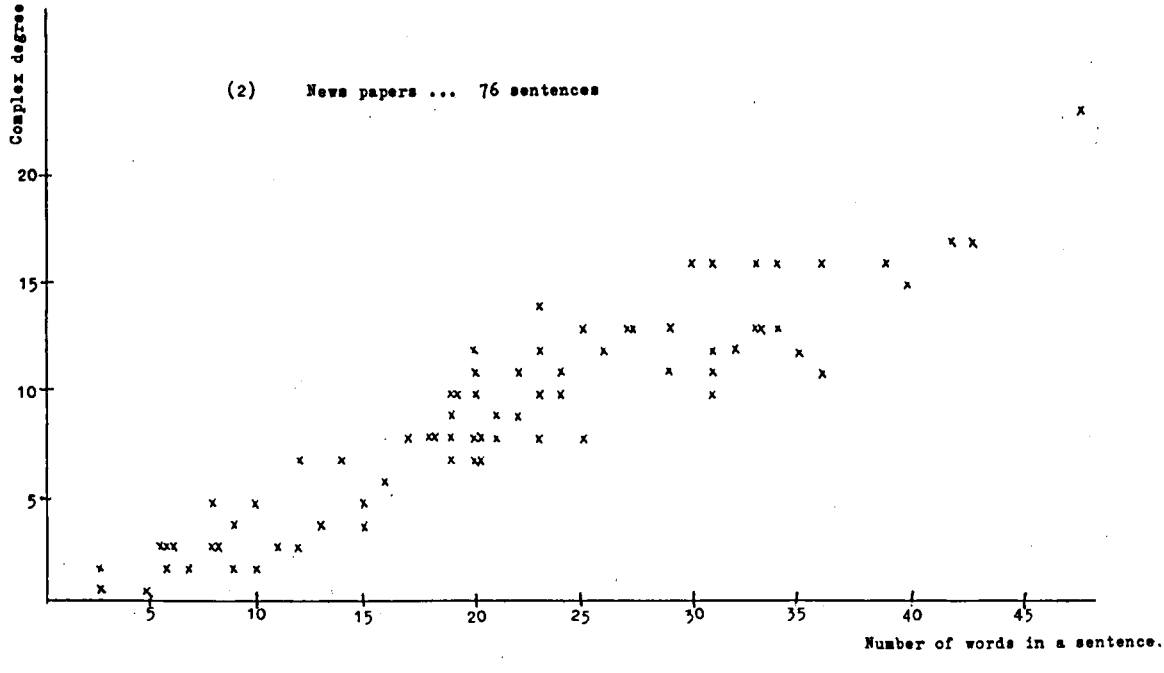
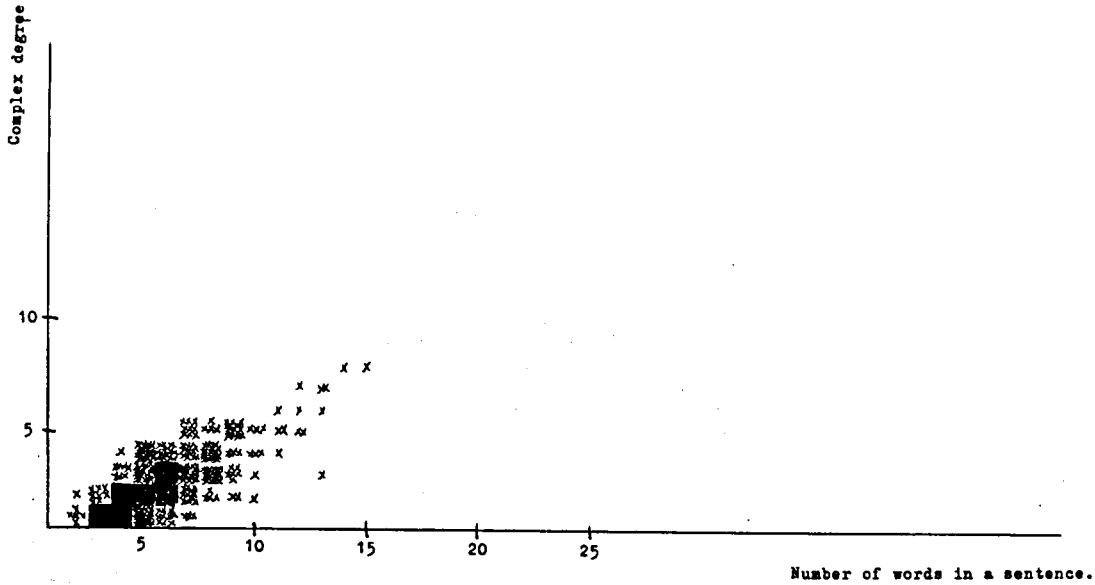


Fig. 6.2.3 Length of the sentence vs complex degree for each sample sentences.

(3) English text .... 762 sentences



(4) Essay ... 52 sentences

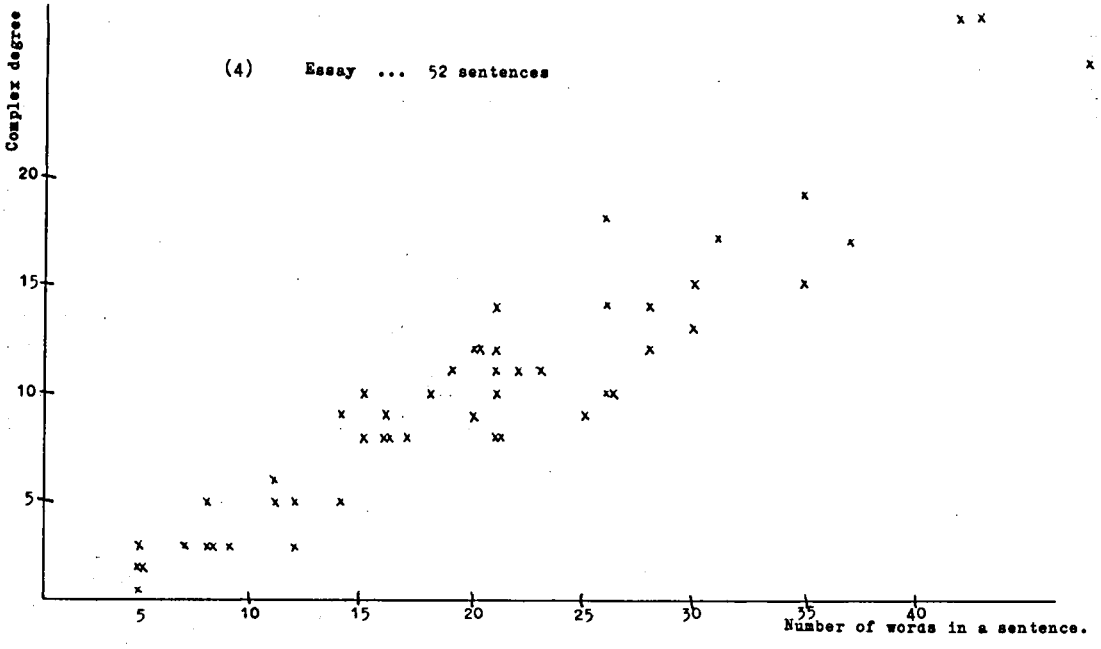


Fig. 6.2.3 Length of the sentence vs complex degree for each sample sentences (continued).

of a sentence, however, can not be determined only the number of function words or by the relative position of those words, but it depends, of course, on the deepness in syntactic structure as well as semantic structure. But the amount of contribution of relative position or deepness to complexity can not be measured objectively only by parts of speech. It is an interesting problem whether it is possible or not to measure the difficulties of a natural sentence as a computable quantity, using only syntactic information.

A connection table of part of speech, which is a frequency list of some sequence of symbols, will become one of the data which characterize the difference between some groups of sentences. Trigrams of part of speech are given in Table 6.2.2.for each field. These tables were constructed by the computer by using almost the same process of translation. That is, input sentences punched on paper tape are fed into the computer from PTR, idiom processing and word dictionary consulting process are carried out, and then parts of speech are given to each word. Then alternation and inference of part of speech is performed. But if the determined parts of speech are not adequate,they are corrected by man. Connection tables are made according to these correct sequence of parts of speech.

The examples, which were used for this experiment, are not enough in number and also in variety. But the purpose of this experiment is not to get a translation score for evaluation of the system, but rather to investigate what kinds of patterns are to appear and how they differ from typical structures, and to find out effective patterns for such structures. Therefore, the score of the result is not so important, but it must be used only as reference.

Table 6.2.2 CONNECTION TABLE OF PARTS OF SPEECH .... 3--SYMBOLS (for Scientific papers)

			FREQ.				FREQ.				FREQ.				FREQ.
NN	PP	NN	0248	NN	PP	DT	0248	DT	NN	PP	0247	DT	NN	NN	0215
DT	AJ	NN	0209	NN	NN	NN	0155	NN	NN	**	0126	PP	NN	NN	0124
NN	NN	PP	0120	AJ	NN	PP	0120	NN	NN	,,	0111	PP	DT	AJ	0105
NN	VL	VP	0103	AJ	NN	NN	0103	PP	AJ	NN	0094	NN	**	DT	0093
**	DT	NN	0085	NN	PP	AJ	0077	NN	,,	C1	0072	((	NN	))	0067
VV	DT	NN	0063	NN	NN	VL	0063	NN	C1	NN	0063	VP	PP	DT	0061
NN	**	PN	0059	VL	VP	PP	0057	PP	NN	PP	0056	NN	,,	NN	0055
AJ	NN	,,	0051	AJ	NN	**	0051	VP	PP	NN	0046	PP	PN	NN	0045
AJ	AJ	NN	0045	PP	NN	**	0042	AJ	NN	VL	0042	,,	DT	NN	0041
NN	VP	PP	0041	--	NN	NN	0041	NN	**	NN	0040	NN	--	NN	0040
DT	NN	VL	0040	P4	DT	NN	0039	NN	PP	PN	0038	NN	((	NN	0037
C1	DT	NN	0037	NN	,,	DT	0036	,,	C1	DT	0033	DT	NN	,,	0033
NN	NN	C1	0032	TO	DT	NN	0031	**	DT	AJ	0031	C1	NN	NN	0031
VL	VP	TO	0030	TO	VV	DT	0030	VV	PP	DT	0029	PP	NN	,,	0029
NN	**	PP	0029	AJ	--	NN	0029	VL	DT	NN	0028	VG	DT	NN	0028
DT	NN	C1	0028	,,	NN	,,	0027	VA	VL	VP	0027	NN	VL	DT	0027
DT	VP	NN	0027	DT	NN	**	0027	VP	NN	NN	0026	VL	VP	**	0025
**	PN	NN	0025	NN	TO	VV	0025	DT	AJ	AJ	0025	NN	AJ	NN	0024
DT	NN	VV	0024	AD	PP	DT	0024	VL	VP	P4	0023	NN	**	AD	0023
NN	NN	VV	0023	VV	NN	PP	0022	**	PN	VL	0022	NN	VV	DT	0022
NN	VL	AD	0022	NN	NN	((	0022	NN	DT	NN	0022	AJ	PP	DT	0022
VL	DT	AJ	0021	NN	**	AJ	0021	NN	RP	VV	0021	VP	TO	VV	0020
VL	AD	AJ	0020	VH	PL	VP	0020	PN	NN	VL	0020	NN	VV	PP	0020
NN	PP	VG	0020	,,	NN	NN	0019	PP	VG	DT	0019	PN	VL	VP	0019

NN=noun, PP=preposition, AJ=adjective,
VV=verb, VP=past verb , AD=adverb,
VG=ing-verb, PN=pronoun, RP=relative pronoun

Table 6.2.2 CONNECTION TABLE OF PARTS OF SPEECH .... 3--SYMBOLS (for News papers)

			FREQ.				FREQ.				FREQ.				FREQ.
DT	NN	PP	0041	DT	NN	NN	0037	PP	DT	NN	0067	NN	PP	DT	0051
DT	AJ	NN	0028	**	DT	NN	0017	NN	PP	NN	0034	NN	NN	NN	0031
NN	NN	,,	0016	NN	NN	**	0016	AJ	NN	PP	0017	PP	NN	NN	0016
VP	DT	NN	0013	NN	VP	PP	0013	NN	NN	PP	0015	VP	PP	DT	0013
DT	NN	,,	0012	NN	**	PN	0011	PP	DT	AJ	0012	NN	**	DT	0012
VP	PP	NN	0010	NN	,,	NN	0010	NN	**	NN	0011	NN	C1	NN	0011
TO	DT	NN	0009	NN	PP	VG	0009	AJ	NN	**	0010	VV	DT	NN	0009
PP	NN	**	0008	NN	VP	DT	0008	NN	NN	C1	0009	AJ	NN	NN	0009
TO	VV	NN	0007	PP	NN	PP	0007	DT	NN	**	0008	VG	DT	NN	0007
NN	NN	VP	0007	C1	DT	NN	0007	NN	TO	VV	0007	NN	PL	VP	0007
NN	,,	AD	0006	NN	TO	NN	0006	PL	VP	TO	0006	NN	,,	DT	0006
NN	C1	DT	0006	DT	NN	VP	0006	NN	TO	DT	0006	NN	))	**	0006
AJ	NN	VP	0006	,,	NN	VP	0005	DT	NN	TO	0006	C1	NN	NN	0006
VP	TO	VV	0005	VG	TO	VV	0005	,,	NN	PP	0005	VV	PP	DT	0005
TH	DT	NN	0005	**	PP	DT	0005	TO	VV	PP	0005	TO	VV	DT	0005
P4	DT	NN	0005	NN	**	PP	0005	PP	NN	,,	0005	PL	VP	PP	0005
NN	AS	DT	0005	DT	NN	))	0005	NN	P4	DT	0005	NN	NN	TO	0005
,,	AD	,,	0004	VV	PP	NN	0004	AJ	NN	TO	0005	,,	DT	NN	0004
VL	VP	TO	0004	VH	VP	NN	0004	VP	TO	VH	0004	VP	NN	NN	0004
**	PN	NN	0004	**	NN	NN	0004	VG	PP	DT	0004	TO	VH	VP	0004
PP	VG	DT	0004	PP	AJ	NN	0004	RA	NN	PP	0004	PP	VG	NN	0004
NN	VV	TO	0004	NN	VP	**	0004	NN	,,	VP	0004	NN	,,	AJ	0004
NN	NN	AD	0004	NN	AJ	NN	0004	NN	"	NN	0004	NN	NN	VV	0004
DT	NN	TH	0004	DT	NN	PN	0004	DT	RA	NN	0004	DT	NN	VV	0004
								DT	NN	C1	0004	C1	AJ	NN	0004

Table 6.2.2 CONNECTION TABLE OF PARTS OF SPEECH .... 3--SYMBOLS (for Text of middle school)

			FREQ.				FREQ.				FREQ.				FREQ.
DT	NN	**	0117	NN	¥¥	**	0083	NN	**	PN	0183	**	PN	VV	0168
PN	VV	NN	0062	**	QQ	,,	0060	**	PN	VL	0075	¥¥	**	QQ	0071
,,	AD	**	0048	VF	PN	VV	0048	VH	DT	NN	0056	PN	VV	PP	0052
**	VF	PN	0047	VV	PP	NN	0044	¥¥	**	PN	0048	QQ	,,	PN	0048
NN	,,	AD	0040	AJ	NN	**	0040	PP	NN	**	0044	PN	VH	DT	0042
AD	**	PN	0036	,,	PN	VF	0035	VV	NN	**	0038	PN	VF	**	0037
VL	DT	NN	0032	PP	DT	NN	0032	**	PN	"	0035	NN	**	NN	0035
PN	VL	NN	0030	NN	"	NN	0029	DT	NN	¥¥	0032	VV	DT	NN	0030
NN	**	VF	0028	**	PN	VF	0027	**	PN	VH	0028	PN	VV	PN	0028
PN	VV	DT	0026	PN	VL	DT	0026	**	QQ	**	0026	PP	AJ	NN	0026
QQ	**	PN	0024	NN	,,	NN	0024	DT	NN	,,	0026	**	DT	NN	0024
AD	AJ	**	0023	VF	PN	VH	0022	PN	VF	VV	0023	PN	"	NN	0023
NN	DT	NN	0022	VV	PP	AJ	0021	**	WW	VF	0022	NN	NN	**	0022
NN	PP	NN	0021	NN	C1	NN	0021	"	NN	**	0021	NN	**	WW	0021
VV	¥¥	**	0019	NN	VV	PP	0019	,,	NN	**	0020	VL	NN	**	0020
PN	VV	AJ	0018	VF	**	PN	0017	WW	VF	PN	0018	**	NN	NN	0018
PP	NN	,,	0016	AJ	**	PN	0016	DT	AJ	NN	0017	**	NN	VV	0016
C1	NN	**	0015	VV	PP	DT	0014	PN	"	DT	0015	NN	**	DT	0015
NN	**	QQ	0014	DT	NN	NN	0014	**	VL	PN	0014	PN	¥¥	**	0014
VL	AD	AJ	0013	**	NN	,,	0013	AD	AD	**	0014	,,	NN	,,	0013
NN	VV	NN	0013	NN	VL	NN	0013	PN	VL	PN	0013	PN	**	PN	0013
DT	NN	VL	0013	VV	AJ	NN	0012	NN	**	C1	0013	NN	PP	DT	0013
**	C1	PN	0012	QQ	,,	NN	0012	VF	**	VF	0012	**	HH	VL	0012
NN	,,	C1	0012	,,	PN	VL	0011	PN	VV	¥¥	0012	PN	VA	VV	0012
								VV	NN	,,	0011	"	DT	NN	0011

Table 6.2.2 CONNECTION TABLE OF PARTS OF SPEECH .... 3--SYMBOLS (for Essay)

			FREQ.				FREQ.				FREQ.				FREQ.
												PP	DT	NN	0045
DT	NN	PP	0022	NN	PP	DT	0021	DT	NN	**	0014	DT	AJ	NN	0014
DT	NN	..	0013	VV	DT	NN	0012	NN	TO	VV	0011	NN	**	PN	0011
NN	PP	VG	0009	AJ	NN	PP	0009	NN	..	C1	0008	DT	NN	TO	0008
PN	VV	PN	0007	DT	NN	AD	0007	AD	PP	DT	0007	VP	DT	NN	0006
TO	VV	PP	0006	TO	VV	DT	0006	**	PN	VV	0006	PP	DT	AJ	0006
NN	PP	NN	0006	DT	NN	C1	0006	VV	TO	VV	0005	VV	PN	**	0005
**	DT	AJ	0005	PN	VV	DT	0005	PN	VA	VV	0005	PN	PP	DT	0005
NN	**	DT	0005	NN	**	C1	0005	NN	C1	DT	0005	DT	NN	NN	0005
C2	PN	VV	0005	C1	VV	PN	0005	C1	DT	NN	0005	..	C1	PN	0004
..	C1	C2	0004	VV	TH	PN	0004	VV	PN	PP	0004	VP	PP	AJ	0004
VL	PP	DT	0004	VG	PP	DT	0004	VG	DT	NN	0004	TO	DT	NN	0004
**	DT	NN	0004	PP	NN	PP	0004	PP	AJ	NN	0004	PN	VV	NN	0004
PN	**	PN	0004	NN	..	PN	0004	NN	AD	PP	0004	DT	NN	VV	0004
DT	AJ	AJ	0004	AD	DT	NN	0004	..	VG	PP	0003	..	PP	DT	0003
..	PN	VV	0003	..	C1	VG	0003	VV	PP	PN	0003	VV	PP	DT	0003
VV	PN	..	0003	VV	PN	AD	0003	VP	PN	PP	0003	TO	VV	PN	0003
TO	VV	AJ	0003	**	PN	VH	0003	PP	VG	DT	0003	PN	..	C1	0003
PN	VP	TO	0003	PN	VL	PP	0003	PN	VH	VP	0003	PN	VF	EE	0003
P4	DT	NN	0003	NN	..	PP	0003	NN	VV	**	0003	NN	P4	DT	0003
NN	C2	PN	0003	DT	NN	VP	0003	C2	PN	VP	0003	C1	C2	PN	0003
AJ	TO	VV	0003	AJ	PP	NN	0003	AJ	NN	VV	0003	AJ	NN	**	0003
AD	..	C1	0003	..	PN	VF	0002	..	PN	VA	0002	..	GH	VP	0002
..	AS	PN	0002	W2	AJ	PN	0002	VV	**	DT	0002	VV	P4	NN	0002
VV	NN	AS	0002	VV	DT	AJ	0002	VV	C2	PN	0002	VV	AJ	C1	0002

6.3 Analysis of results

The results of mechanical translation for the samples shown in section 6.2 are roughly scored in Table 6.3.1, which were obtained by using the rewriting rules listed in Table 3.5.1 in section 3.5.

Table 6.3.1 Score of the results.

	Correct	Almost correct	Erroneous
(1) Scientific papers	58 %	23 %	19 %
(2) News papers	45 %	23 %	32 %
(3) Text book of middle school	96 %	2 %	2 %
(4) Essay	41 %	23 %	36 %

The correct results are such ones whose tree structures are perfect as shown in Fig.6.3.1. A sample of "almost correct results" is shown in Fig.6.3.2, which has not a complete tree, but its main part is correctly analyzed, or a slight modification of input form leads to a correct analysis. For example, an original sentence "THE PRESENT PROGRAM DEALS ONLY WITH AND AND NOT TYPE SEARCHES" is not analyzed correctly because of a irregular expression of "AND" and "NOT", if the first "AND" and "NOT" are changed to"AND*" and "NOT*", they will be regarded to be nouns, and the sentence is correctly analyzed.

Examples of erroneous cases are shown in Fig.6.3.3 or Fig.6.3.4, in which several sub-strings are parsed correctly, but as a whole their mutual relation are not analyzed correctly.

An above mentioned evaluation is very rough, and even inadequate to investigate the degree of fitness of phrase structure grammar to the syntax of English. In order to examine the adequacy of phrase structure grammar, more detail analysis of results must be adopted. Such a method

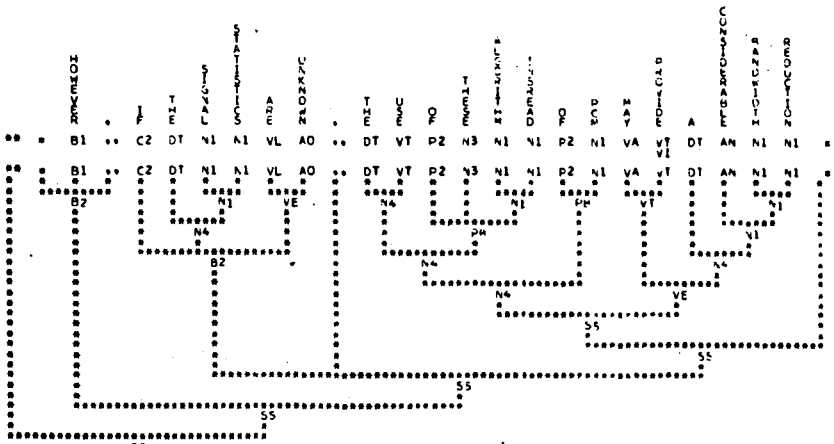


Fig. 6.3.1
An example whose syntax is analyzed correctly.

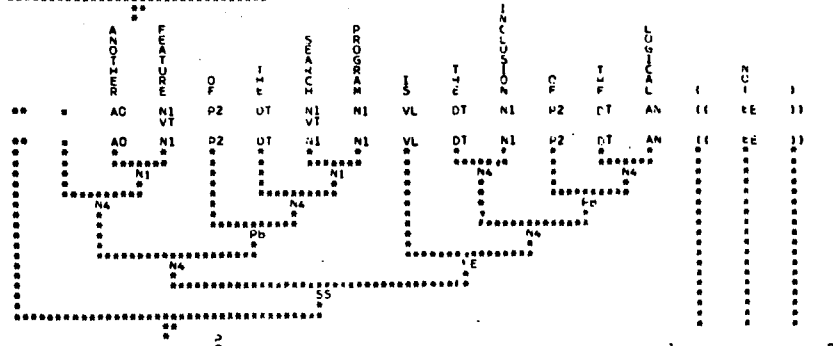


Fig. 6.3.2
An example whose syntax is analyzed almost correctly.

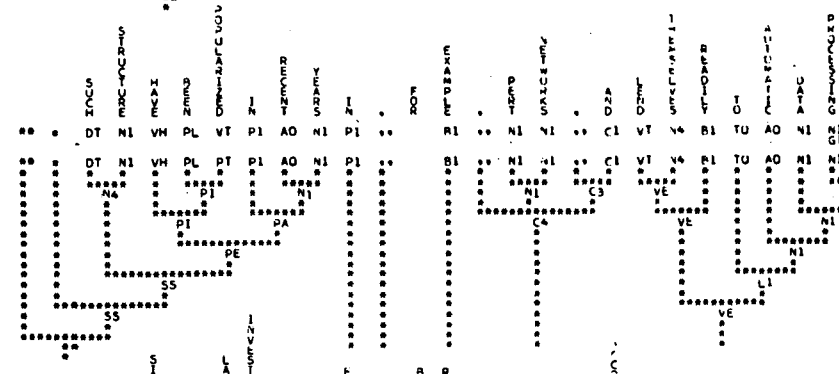


Fig. 6.3.3
An example which is not analyzed correctly.

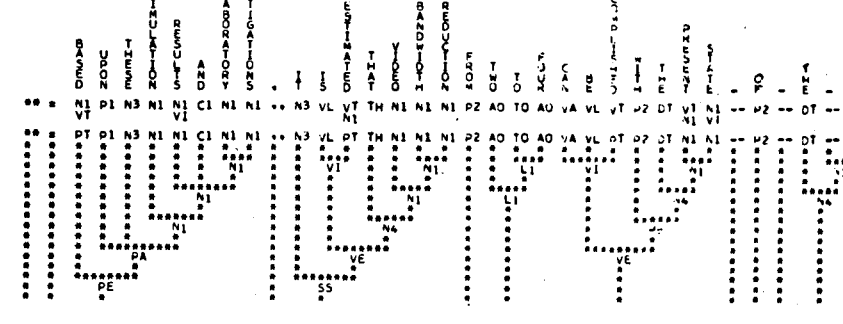


Fig. 6.3.4
An example which is not analyzed correctly.

as mentioned in the following may be one evaluation.

That is, in the structure tree of English which was printed by the computer, erroneous branches are cut off and only correctly parsed branches are left, and the number of independent sub-tree are counted. This number will be called, "resolution degree". Then the mark of a result is equal to complexity degree minus resolution degree plus one. For example, Fig.6.3.5-a is a result of translation, but checked branches are wrong parsings and cut off, so that the amended tree becomes like Fig.6.3.5-b, which has 5 sub-trees (a beginning mark and a boundary mark are ignored). Therefore, the mark of the sentence in Fig.6.3.5 is 1 (= 5 - 5 + 1).

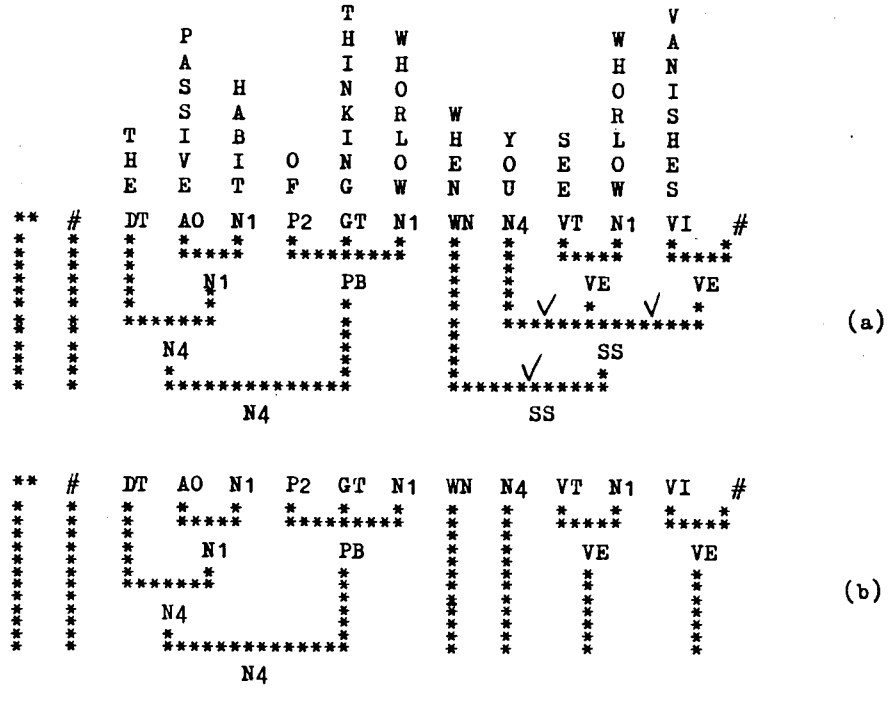


Fig. 6.3.5 Calculation of "resolusion degree" : (a) result obtained by MT, (b) decomposed form by cutting off erroneous branches.

By this evaluation, the adequacy of phrase structure grammar becomes about eighty-percent for scientific papers.

Such kind of marks mentioned above may not be effective to evaluate the translation system from the practical point of view, but they are useful as reference data to understand how much the phrase structure grammar can be applicable to the English sentences. If such a result as Fig.6.3.2 is classified into erroneous groups only because it has not a complete tree, it may be too severe evaluation to adopt as an adquate judgement of the system.

The above mentioned evaluation treats only the general aspects of translation results, but in the following, erroneous results are examined one by one, and difficult points in mechanical translation from English into Japanese are investigated. Furthermore, what kinds of improvement in algorithm and rewriting rules, or restriction to input sentences can make them correct is considered, and this leads to the pre-editing and post-editing. The meaning of "correct", however, is only in syntactic level, and semantic aspect is not considered.

In the gross, erroneous results are classified into the following four cases.

(1) Errors by mis-determination of part of speech(55 %)

(2) Errors by lack of rewriting rules (20 %)

(3) Errors by wrong rules or wrong hierarchy(15 %)

(4) Errors by irregular forms of input sentences(10 %)

Percent numerals in parenthesis indicate the error ratio to total erroneous results. They are not rigid numbers but only for reference. Each case is explained below.

6.3.1 Errors by mis-determination of part of speech

There are several instances in errors of part of speech. One of them is the case where classification of part of speech is inadequate. For example, the words "half" works as noun or adjective or prepositional word and "above" works as adverb or preposition. Nouns, adjectives and adverbs belong to form class, and prepositions to function class. In the present system, a function word can not have two parts of speech; therefore a new part of speech must be given to such a word. "Equivalent" or "oneself" etc. also plays an irregular role as compared with an ordinary adjective or pronoun or adverb, so a characteristic symbol must be given to these words. In the experiment "half"(in Fig.6.3.1.1), "above"(in Fig.6.3.1.2), "equivalent" and "oneself" are named N1, P1, AN, and N4, but they are inadequate. Next, idiomatic expressions are considered. The parts of speech of idiomatic expressions are sometimes difficult to determine because of their rather irregular use. For example, in the first sentence, shown below, the idiomatic expression " , that is, " is used as a conjunctive phrase which connects words or phrases, but sometimes the same expression connects clauses, as shown in the third sentence.

(1) " Perhaps the most difficult problems are associated with syntax, that is, structure of phrases and sentences "

(2) " It is certainly desirable to have document indexes , that is, descriptor lists, as short as possible "

(3) " The principal rule does not change, that is, the nearer they are to the main verb, the more intimate they are to it. "

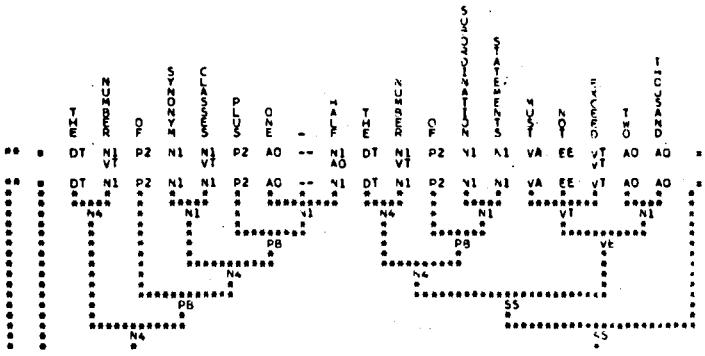


Fig. 6.3.1.1
An example which is not completely analyzed because of wrong part of speech given to the word "half".

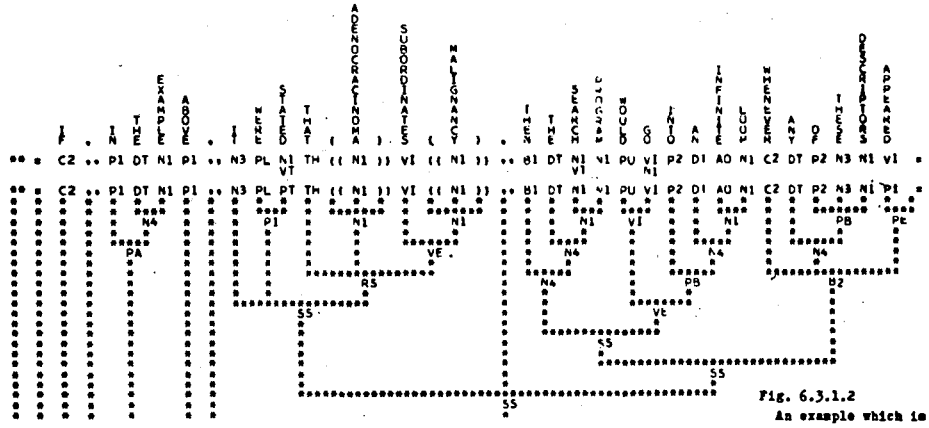


Fig. 6.3.1.2
An example which is almost correct if the part of speech given to "above" is adequate.

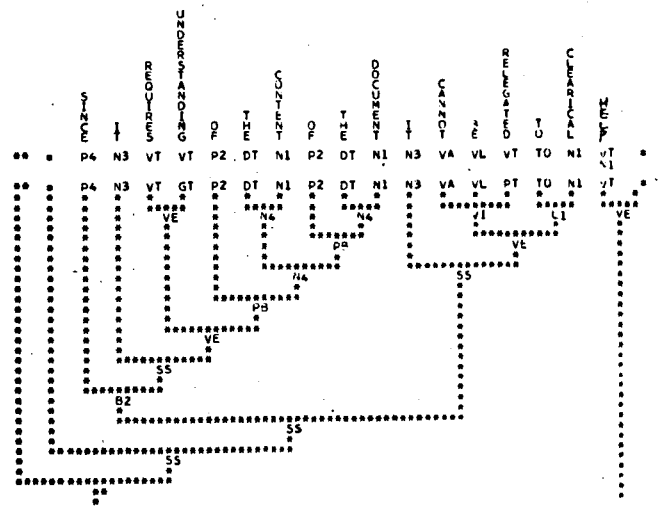


Fig. 6.3.1.3
An example whose tree is not completed because the selection of part of speech out of two possibilities is erroneous for the word "help".

Therefore, if C1 symbol (which means co-ordinate conjunction) is given to " , that is, ", sometimes wrong parsing is made for the third example.

Errors which appear most frequently in part of speech are failures to select an adequate one out of the two symbols which are given to a word. For example, in Fig.6.3.1.3, the last word "help" which has two parts of speech VT and N1 in the dictionary was looked on as VT when syntactic analysis was carried out. Because no key words such as determiner or preposition or auxiliary verb can be found just before the word "help", the first part of speech, which is usually used more frequently than the second one, is adopted. The order between two parts of speech given to a word in the dictionary is the same as that of an ordinary word dictionary as a rule. If, in Fig.6.3.1.3, the word on the left side of "help" is a key word, its part of speech is correctly determined as noun by the algorithm mentioned in section 5.1.

In Fig.6.3.1.4, "implemented" and "study" are wrong in part of speech. As regards "implemented", only the noun use is registered in the wrong dictionary so the word ending "ed" is ignored and "implemented" is looked upon as a noun. If it is not registered in the dictionary, the ending "ed" is taken into consideration and the part of speech of "implemented" is infered as PT (past form verb). This error is caused by the defect of word dictionary because the verb use of "implement" is overlooked, the same holds with a word "subordinate" in Fig.6.3.1.5. The case of "study" is the same as the previous example "help". If, in these cases, the context of wider range is taken into consideration, it may be thought possible to choose an adequate part of speech, but it is difficult to determine part of speech from the static sequence of part of speech, "static" here meaning that each word is given a part of speech independently of other words. If dynamic sequence must be con-

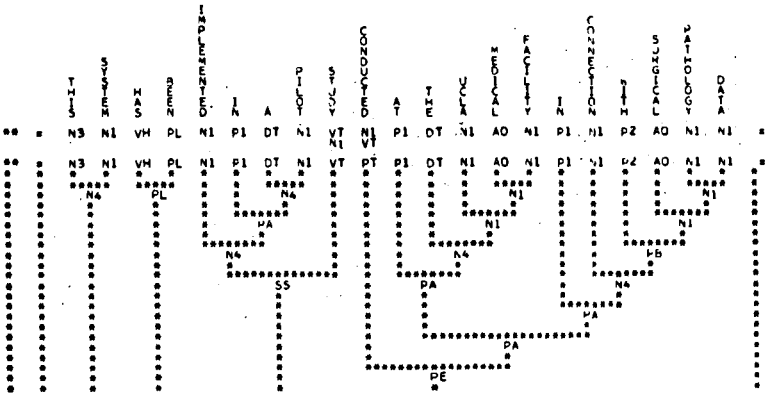


Fig. 6.3.1.4
An example which is not analyzed correctly because of the wrong parts of speech for "implemented" and "study": as for the case of "implemented", it is caused by the defect of word dictionary. As for the case of "study", the context is not enough to select "N1" from "VT" and "N1" given to "study".

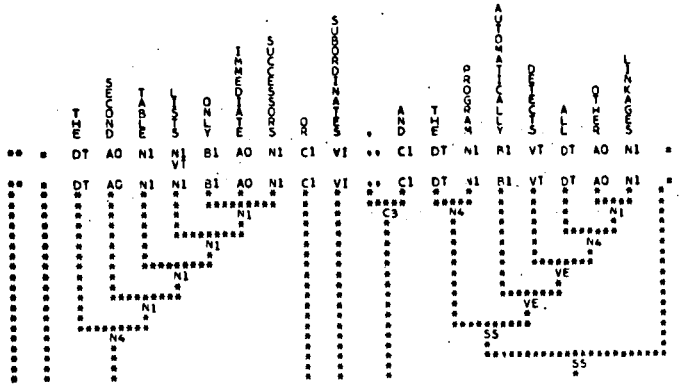


Fig. 6.3.1.5
An example in which parts of speech for "list" and "subordinates" are not adequate: as for the case of "subordinates", it is caused by the defect of word-dictionary.

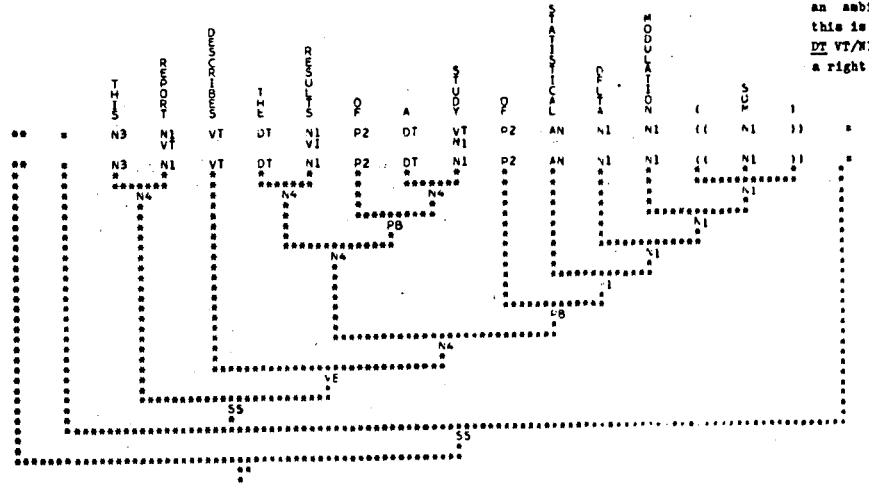
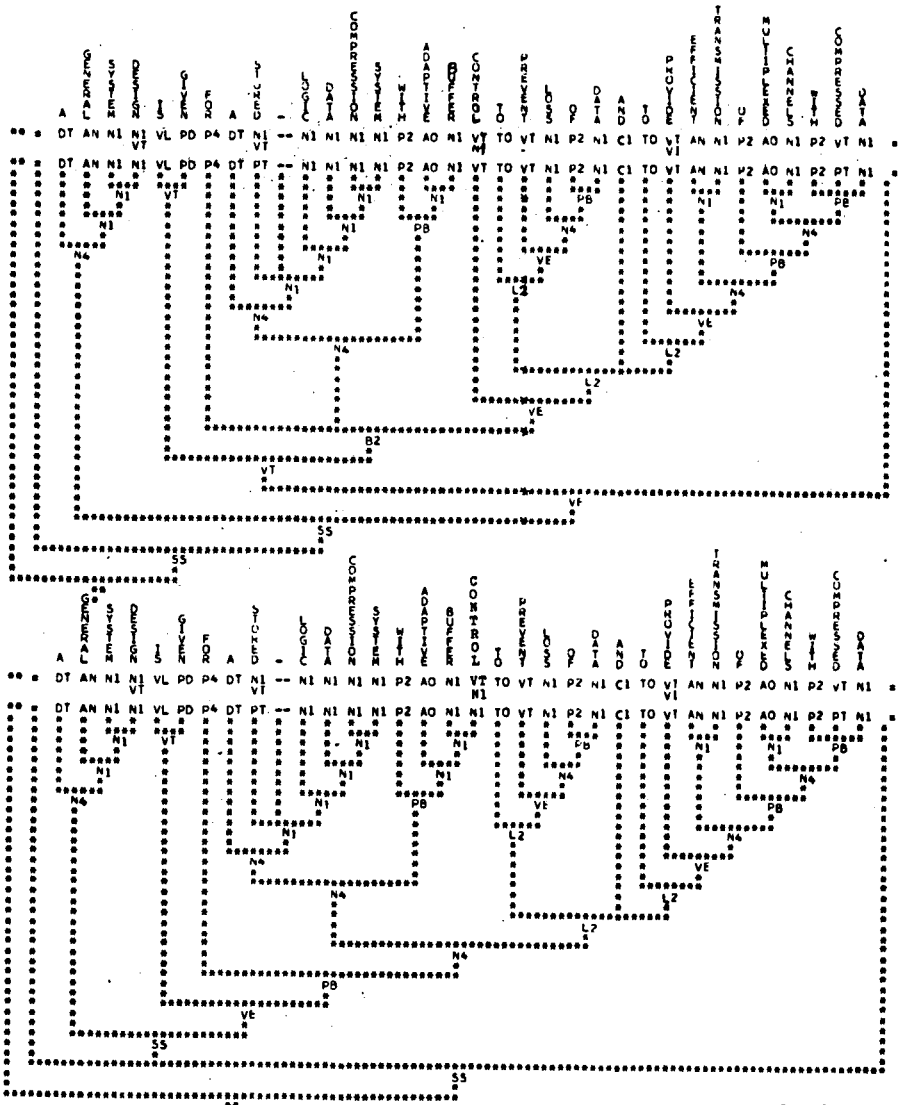


Fig. 6.3.1.6
An example in which a correct part of speech is selected for an ambiguous word "study": this is because the context of DT VT/N1... is enough to select a right one "N1".

sidered, however, the analysis of sentence must be carried out, but to do so part of speech must be determined. It is a contradiction. This is one of the most important problems to be solved in machine translation. In the case of "study" in Fig.6.3.1.6, it is correctly chosen as a noun because it appears just after a determiner.

Analysis of all possible combinations of parts of speech for the words in the sentence is recommended, if there are several words which have two parts of speech and can not be determined uniquely by the context or word-endings. Contrary to expectation, however, the numbers of possible combinations is comparatively small, but the solution of the problem is far from complete. Because there exist many cases where even the wrong sequence of parts of speech can be analyzed into the complete tree. For example, in Fig.6.3.1.7, the word "control" is looked upon as VT (present form verb) in stead of a right one (N1), but the structure tree is completed, so that there is a fear that even a wrong part of speech is looked upon as a right one. A right tree is of course obtained by giving a right part of speech N1 to "control". Though their tree structures are different from each other, it can not be determined which of them is correct from the syntactic point of view, but it depends on the semantic context of the text. Then, it is necessary to know what kind of semantic information will determine the part of speech of, for example, "control" in Fig.6.3.1.7, but it is almost impossible in the present state of linguistic knowledge in the computer. Therefore, we must seek more effective algorithm to select adequate parts of speech by making the most use of syntactic information.

In some cases, word-end processing causes a confusion in giving a part of speech to a word. A word which is not actually registered in the dictionary is regarded as if it were found in the dictionary,



(a)

(b)

Fig. 6.3.1.7
An examples which are both correctly analyzed; in (a) is looked on as a verb, in (b) the same word is looked on as a noun.

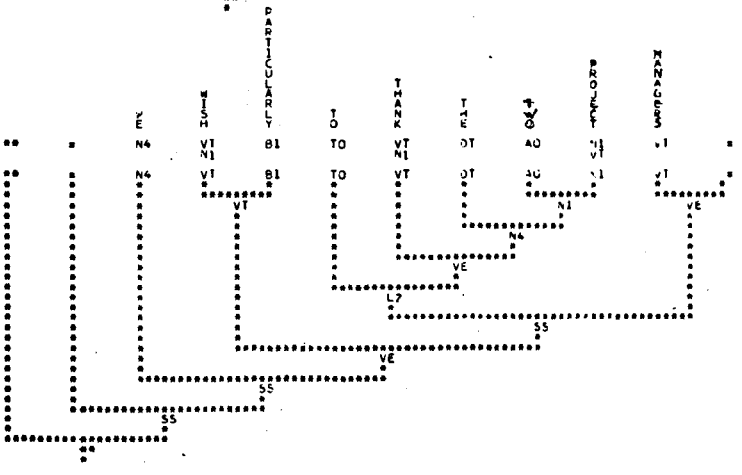


Fig. 6.3.1.8
An example which is erroneous because of the defect in the word dictionary, that is, "managers" is looked on as the same with "manage" (VT) as a result of word-end processing.

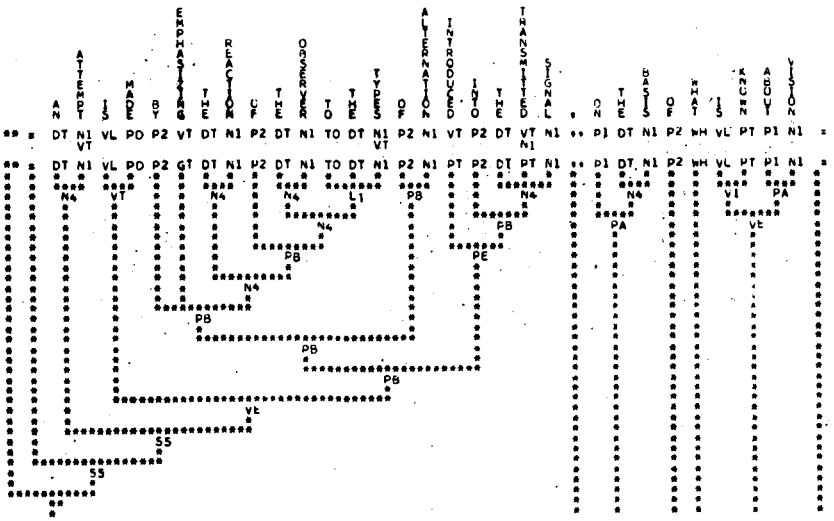


Fig. 6.3.2.1
An example whose syntax is somewhat complex.

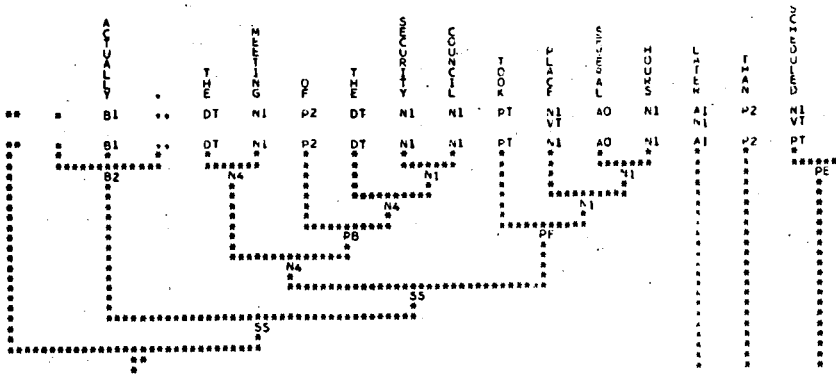


Fig. 6.3.2.2
An example which includes a rather irregular expression "later than scheduled".

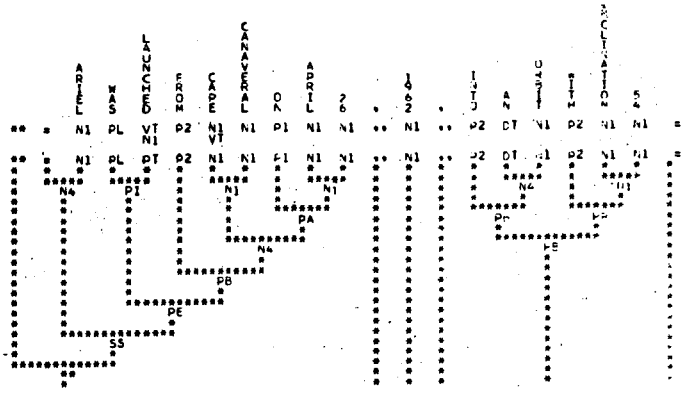


Fig. 6.3.2.3
An example which includes date expression which is left unanalyzed.

because its pseudo-stem becomes the same with that of other word. For example, in Fig.6.3.1.8 a word "managers" is given a part of speech " VT ", which corresponds to that of "manage ". This is because pseudo-stems of "managers" and " manage " are the same, and an entry corresponding to "manager" is not registered in the dictionary, therefore the word "managers"is regarded to be coincided with "manage " by the algorithm mentioned in section 5.1.3. This error, however, is easily solved by registering a word "manager" into the dictionary in a form shown below.

MANAG	VT	KANRIS
ØØØER	N1	KANRISYA
ØØERS	N1	KANRISYA

6.3.2 Errors by lack of rewriting rules

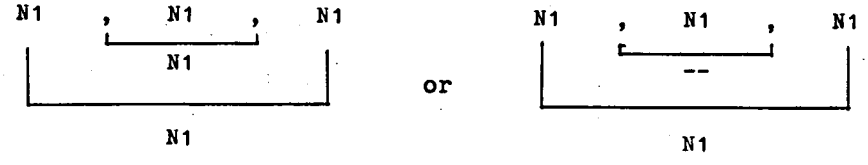
The cases where rewriting rules do not exist occur when irregular or unfamiliar expressions appear which are not previously expected. In these cases, it is generally difficult to make rewriting rules. For example,in Fig.6.3.2.1,the later part of the sentence "on the basis of what is known about vision" remains unanalyzed, because it is doubtful whether it is possible to make such a rule as P2 SS PB which means preposition takes a clause as its object. If such rules are added newly, it may interfere other normal cases. Therefore the introduction of new rules must be carefully investigated by analyzing many examples.

Other examples which contain rather an exceptional sequence of parts of speech are shown in Fig.6.3.2.2. The sequence "RA P2 PT" which corresponds to "later than scheduled" may be classified into an incomplete case of the original sentence, but it is desirable to be treated as it is by the rewriting rules because such an expression

appear comparatively frequently, though the forms of the rewriting rules which treat these cases can not be definitely determined.

Further, examples which contain "as" are tedious and difficult to be processed by the fixed context-free type patterns, although "as" is distinguished from all other words, and their rules are used according to their hierarchy. Especially in the expression "as A as B", various kinds of words, phrases, or sentences can come in the position A or B. A rule whose length is limited to less than three makes it difficult, in this case, to seize enough context to determine the role of phrases. That is, possible combinations are "A as B", "as A as", "as A", or "as B", and their substitution symbol differs from case to case, so that their rule can not be uniformly determined.

Another case belongs to a special one but appears very frequently in scientific papers. It is the date expression as shown in Fig.6.3.2.3. Numerals are not distinguished from noun class, and they are included in ordinary noun class, so it is difficult to descriminate the date expression from the appearance of " , Nl, ". Then a parsing given below is generally not a suitable one because it lessens the information which commas convey, though convenient in this case.



If, however, the limitation of kinds of input sentences guarantees that sentences which include nominatives of address do not appear, the pattern ",,Nl,, Nl" is effective because this pattern can be applied

to the case of aposition or array of nouns. But if there seems to be some apprehension of confusing cases, it is the most sound solution to deform the input sentence itself by the pre-edition, about which will be explained in the next section, without introducing the new rule ",,N1,, ⟶ N1".

Furthermore, it is difficult to make adequate rules when commas or hyphens are inserted to make the sentence easy for man to read, because such special symbols separate the originally adjacent words or phrases, so the rules of restricted length can not sometimes grasp correctly the mutual relation between words or phrases squeezed by a comma. For example, in Fig.6.3.2.4-a, the existence of comma makes the structure complex. If the comma is rejected, its structure becomes a typical form, so the correct parsing is performed as shown in Fig. 6.3.2.4-b.

The existence of commas or hyphens is very useful for man to understand easily or smoothly when long phrases or clauses are intermingled. As for the mechanical processing, however, it has nothing to do with the length of phrases or clauses nor with the formal complexity if the structure obeys the typical grammar. In other words, the existence of commas makes the sentence readable for man, rather than it gives him grammatical information. He can generally understand without comma. To the machine, however, commas give an important clue to the analysis of sentence structures. Therefore their existence have a great influence on sentence structure.

The same discussion holds with expressions which consists of an array of nouns or verb phrases. A typical form in arranging several things is "A, B, C, and D". But in actual sentences there are various kinds of deformations, such as "A, B, C, D", "A, and B, and C, and D"

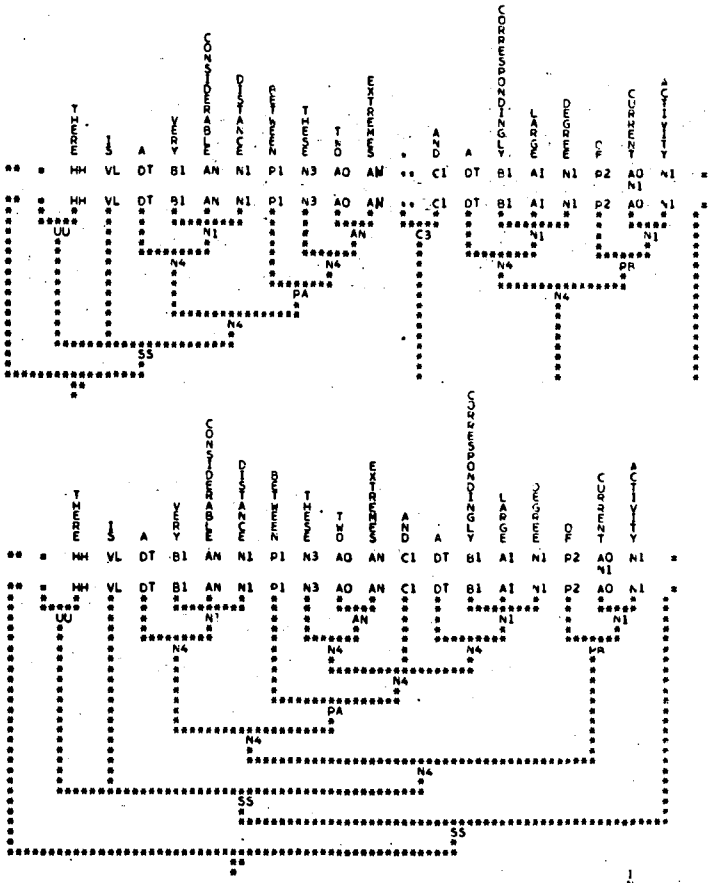


Fig. 6.3.2.4
(a) An example which is not correctly analyzed because of abuse of a comma; (b) an example which is correctly analyzed by deleting the comma in "a).

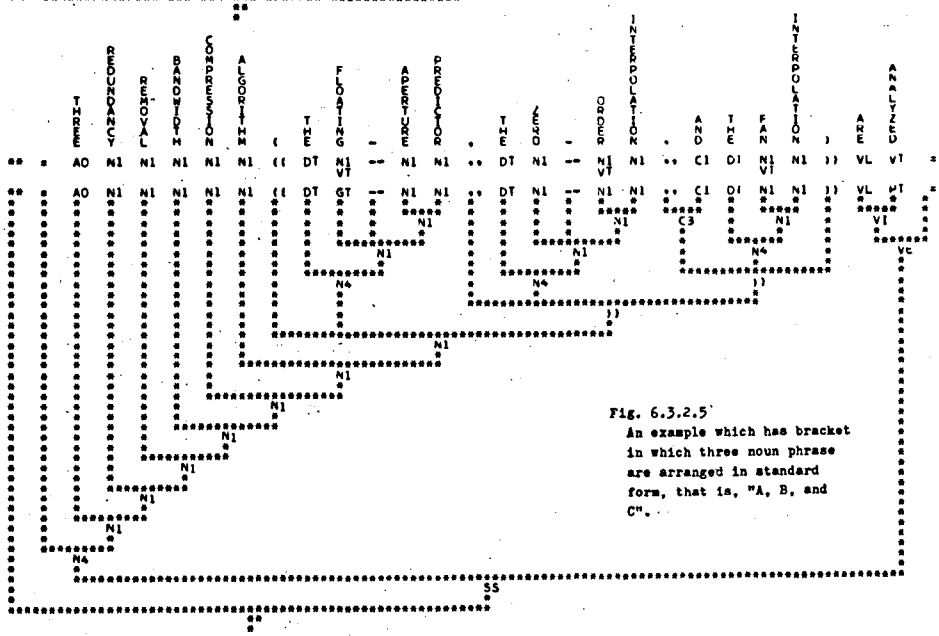
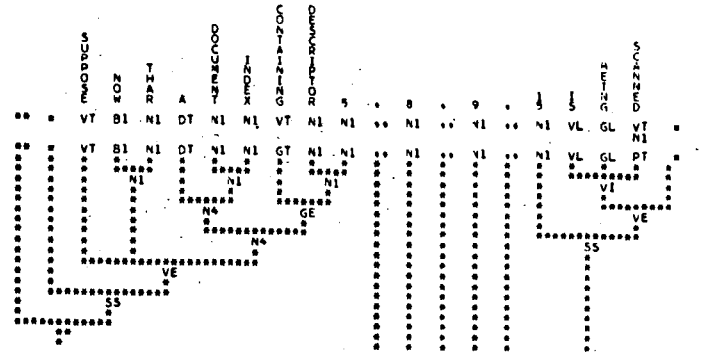
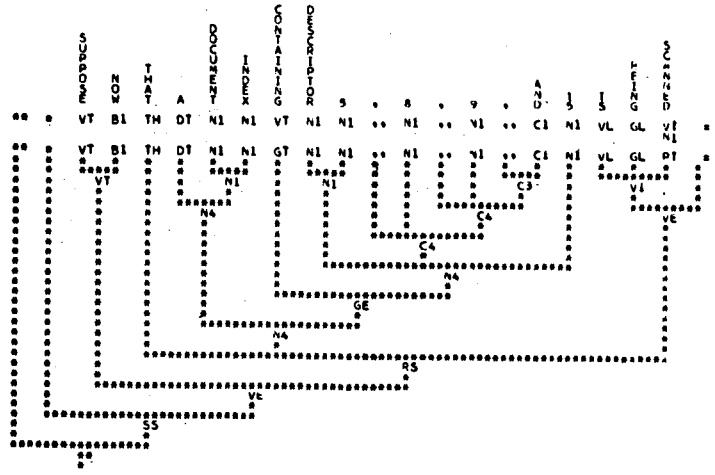


Fig. 6.3.2.5
An example which has bracket in which three noun phrase are arranged in standard form, that is, "A, B, and C".



(a)

1001



(b)

Fig. 6.3.2.6
(a) An example which includes an irregular expression in arranging nouns: (b) an example which is analyzed correctly by inserting a conjunction "and" between the third comma and numeral "15" in (a).

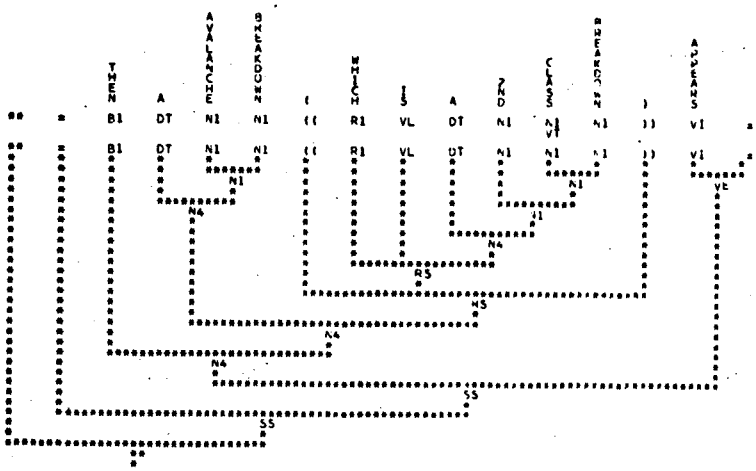


Fig. 6.3.2.7
An example which shows a relative clause in brackets is analyzed correctly.

or "A, B, C and D". For example, in Fig.6.3.2.5, the inner part of parenthesis is a typical form "A, B, and C". But in Fig.6.3.2.6-a, a "and" is missing, therefore the array expression is left unanalyzed. Insertion of "and" leads to a correct analysis as shown in Fig.6.3.2.6-b.

As regards the use of parenthesis, generally the part of speech of the words or phrases set in the parenthesis is the same as that of the words or phrases before the parenthesis as shown in Fig.6.3.2.5. Sometimes parentnesis are used as a supplementary expression, in such case as shown in Fig.6.3.2.7 it is only enough to treat them by the rule "((·RS·)) ⟶ RS" and "((·PA·)) ⟶ PA". But if parenthesis expressions require translating into Japanese in the same form as English has, sometimes somewhat adhoc rewriting rules must be introduced. If, however, parenthesis need not be expressed in Japanese, the simplest solution is to ignore them when they are fed into computer.

6.3.3 Errors by wrong rules or wrong hierarchy

There are some rules which are not necessarily applicable to all cases. This is a kind of ambiguity in syntax. Such rules are effective in some cases, but cause erroneous analysis in other cases. For example, the rule "N4·N4·VE ⟶ N4 (2 3 1 , GA·Ø·Ø)", which is constructed by supposing the actual case where a relative pronoun is abbreviated, works well in the sentence "The range the samples are expected to occupy within a block is predicted from observation of the previous block.", but in such a sentence as "For this reason the program has been written to accomodate batched-requests.", the same rule leads to an erroneous result. Also the rule "N4 ,, SS ⟶ SS (1·2·3 , ØØØ)"

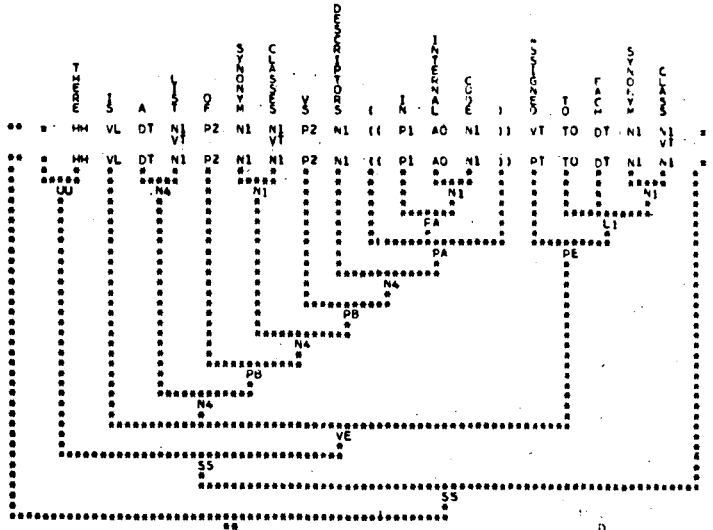


Fig. 6.3.2.8
An example which shows a prepositional phrase in brackets is analyzed correctly.

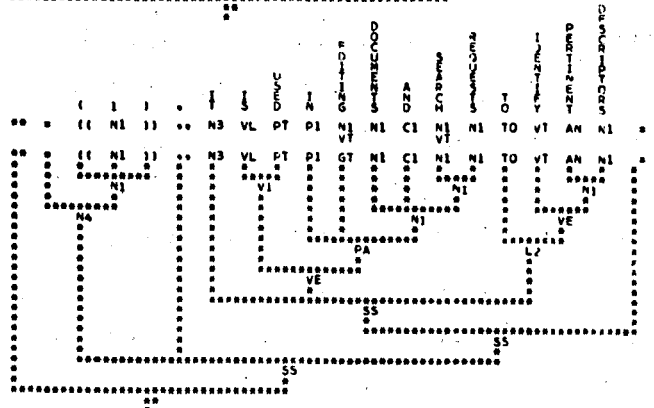


Fig. 6.3.3.1
An example which includes brackets.

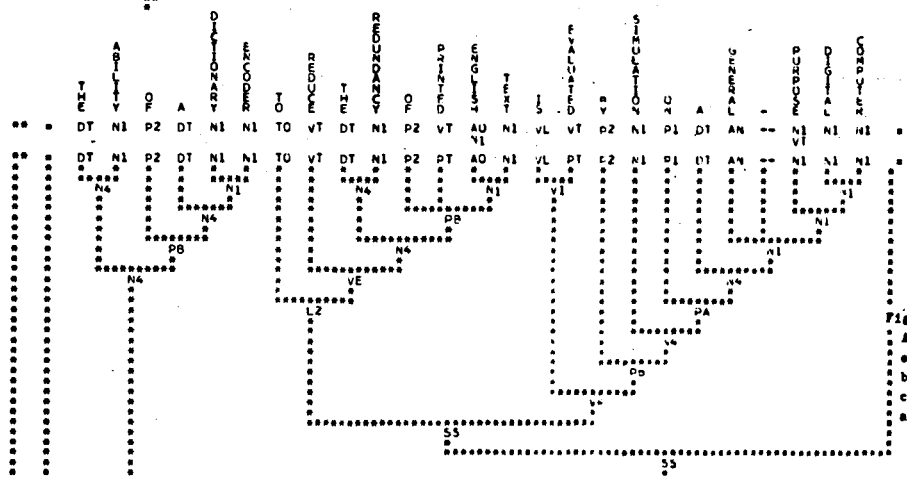


Fig. 6.3.3.2
An example which is erroneously parsed by the wrong hierarchy for N4·L2→N4 and L2·VE→SS.

treats very well such cases as "(1), it is used in editing documents and search requests to identify pertinent discriptors." (Fig.6.3.3.1), but if such a sequence of symbol appears in the middle part of a comparatively long sentence, the noun N4 may be an object of some verb, therefore this rule also has the possibility of giving a wrong analysis.

In such a sentence as this, "In order to make change in logic structure, including those changes which are required to create the file initially, a set of requests is provided", its structure becomes "L2, GE, SS" in the course of analysis. Then, the adoption of one of two rules ",, GE ,, ⟶GE (2 Ø Ø , Ø Ø Ø)" (adjectival phrases which modifies nouns from the right side) and " ,, GE ,,⟶B2 (2 Ø Ø , ITUTU Ø Ø)" (adverbial phrases) affects the following analysis. In real sentences both cases occur, and their selection depends on the semantic information, but not the syntactic one, which the phrase GE conveys in the sentence. In the present system, however, patterns are prohibited from having the same head (the left side of the arrow in the rewriting rule), so the above two rules can not be co-exist, and only " ,, GE ,, ⟶ GE" rule is introduced because this rule is perhaps more probable than another one, if we reflect our activity of generating sentences which have such structure.

Although in the next erroneous case the rules themselves are not ambiguous, but their application order (hierarchy) is wrong. As shown in the middle part of the analysis tree in Fig.6.3.3.2, "N4 L2 VE" is parsed wrong because of the two rules "N4 L2 ⟶ N4 (2 1 , TAMENO·Ø)" and "L2·VE ⟶SS (1·2 , KOTOWA·Ø)" which are in the same class the later rule is applied first by the right-to-left parsing method. But strictly speaking, the former rule must be applied before the latter one, namely, "N4 L2 ⟶N4" is to be stored in A class, and "L2 VE ⟶

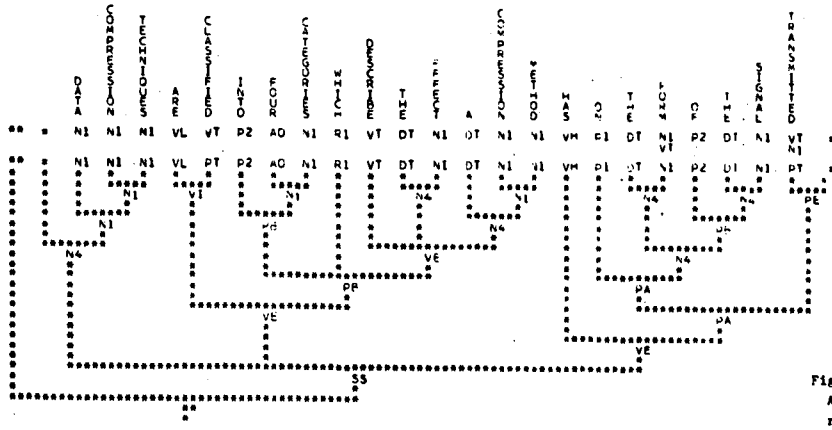


Fig. 6.3.4.1
An example in which a relative pronoun is missing between "effect" and "a".

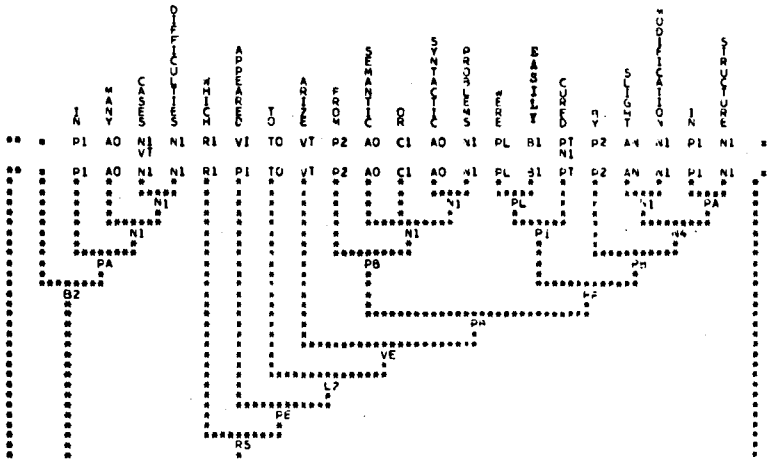


Fig. 6.3.4.2
An example which is parsed wrong because of the absence of a comma between "cases" and "difficulties", and also because of the wrong pattern "PB·PE → PB".

SS" is B class. There are a few other rules which are stored in a wrong class. They must be revised through the actual translation of many kinds of sample sentences.

6.3.4 Errors by irregular form of input sentences

When input sentences are difective, there are, as a matter of course, possibilities of errors. The majority cases of defection in input sentences are those of omission of some words. For example, a complete form of the sentence shown in Fig.6.3.4.1, which apparently seems to be correct in analysis though it is erroneous, must be as below.

" Data compression techniques are classified into four categories which describe the effect which a compression method has on the form of the signal transmitted."

In this case, it is almost impossible to judge whether there is an abbreviated word or not from the syntactic point of view. Another example in which a verb phrase is omitted is shown below, the verb phrase in brackets being omitted. Therefore, the underlined part is looked upon as the predicate for the second sentence.

" Experience with the language of surgical pathology is given, also an out put line of an automatic data processing system applied to the problem [is given] ."

The omission of verbs is rather less frequent than that of relative pronouns. For example,

" TAIHO, himself, replying to reporter, said (that) he did not think (that) his marriage would affect his wrestling. "

" Simple calculation shows that the minimum distance (which) the fission fragments would have had to travel in the maximum permissible time of 30 sec is about 5600 km ."

These sentences appear so natural that we can not think there are omission of relative pronouns, but in the case of a mechanical processing they are looked upon as quite different structures according as there exist relative pronouns or not.

In some cases the boundaries of phrase or clauses can not be clearly recognized by lack of commas. For example, in the next sentence the right most word in the prepositional phrase at the beginning of the

" In many cases difficulties which appeared to arize from semantic or syntactic problems were rather easily curred by slight modification in structure. "

sentence is looked upon as "difficulties", because both "cases" and "difficulties" have N1 as part of speech, and there is no comma between them, so "cases" is looked upon as if it modified "difficulties" by the existence of the rule "N1·N1 ⟶ N1". Also in the same sentence above, the predicate part led by "were" is analyzed as if it modified the prepositional phrase led by "from" by the rule "PB·PE ⟶ PB (2·1 , RARETA·0J", as shown in Fig.6.3.4.2. But this error can be avoided by the insertion of a comma just before the word "were", though the error is caused by the fact that the two rules "VT·PB ⟶ VE" and

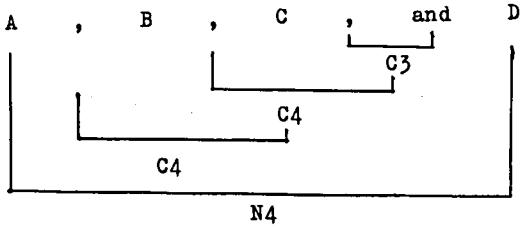
"PB PE ⟶PB" are set in the same class. The next sentence is also the same case as the above example. A comma must be inserted just after the word "problem".

" In the present problem ^ association memory has been provided."
,

On the contrary, an extra comma sometimes makes the structure too complex to be recognized by fixed length patterns. For example, in the sentence

" In order to fulfill these requirements it must relate descriptors in their prose form to internal codes (,) and also to a synonym class."

a comma just before the word "and" isolates the word group following "and", therefore it must not exist. This is caused by the fact that the rule "N4·C3·N4 ⟶N4 (1·2·3 , Ø·Ø·Ø)" can not exist, because this rule, if it exists, leads the sequence "N4 VH DT N1 ,, C1 N4 VH DT N4" which corresponds to "I have a book, and he has a book" to the erroneous parsing "N4 VH N4 VE" which corresponds to "I have (a book, and he) has a book". Words or phrases which are to be connected together must be connected directly by the co-ordinate conjunction (C1) like A C1 A, or by the use of " ,, C1 ⟶C3" (A class), " ,, N4·C3 ⟶ C4" (A class), and "N4 C4 N4 ⟶ N4" (B class) as below.



p205,206 are misplaced after p208.

can be clearly recognized, so special signs are not necessary.

" It is written twice on an output tape, where its association with request is noted for later sorting. "

Among the adverbial uses, if the past-form verb modifies the noun from the right side, such a verb has the sign *, attached to it.

" Another function of the editted program is to *produce hard copy, *giving the report itself as key-punched from original document together with the lists of description *extracted from text."

Underlined word is used to modify nouns from the left side, and so * sign is not attached to it (but it is not error to attach sign). As for "giving", special sign * is not necessarily needed, because the next word is a determiner which is not modified from the left side.

Although this pre-edition is useful, but it is rather doubtful whether this is an easy process or not. In some case where text is complex or highly special, it may be probably difficult to put sign to verb without understanding the content of the text, but generally this processing is easy for everyone.

(2) Substitution of ambiguous words.

There are several words which have many functions in similar contexts but do not appear so frequently. As regard such words one solution is to substitute them with words fit for the context. For example, "about" works both as an adjective and a preposition, and it is often difficult to distinguish its function from the limited context. Therefore, when "about" is used as an adjective, it is enough to substitute "about" with another word, say "almost". For example, in the

example below, "about" is substituted by "almost".

" The dictionary for surgical pathology developed at UCLA presently includes about several thousand descriptors. "
almost

As for several words which have two parts of speech, that is, conjunction and preposition, sometimes their function can be distinguished by the context, but it is not always possible to distinguish a preposition from a conjunction, especially such words as "for", "after", "before" etc. Then, such words must be substituted by other words when they are used as a conjunction.

Further, there are several other words whose function it is difficult to determine because they are used in a quite different way. Such words are "need", "may", "half", "near", or "close", "like" etc. "May" is used as an auxiliary verb and also as a name of the fifth month; therefore, when it is used as the latter meaning, something must be done to distinguish from the former one.

In some expressions, abbreviated words or signs are confusable with ordinary words because capital letters and small letters are not distinguished.* For example, in the example "Fig A and I show...", A is confusable with a determiner, and sign I is confusable with a personal pronoun. The sequence of parts of speech for an above phrase becomes "N1 DT C1 N4 VT ...", and this leads to quite a nonsensical result. Therefore, abbreviated words such as MISS, US, etc. and signs must be adequately transformed by, for example, attaching a special letter to them, before they are fed into the computer.

(3) Insertion or deletion of comma and hyphen.

As mentioned in previous section, the absence of comma makes the boundary of phrases vague, whereas the excess of comma and hyphen makes

---

* It is easily distinguished, if additional information is added.

A symbol C3 (= ,, + C1) is used to connect other phrases than noun equivalents. For the same reason the second comma in the next sentence must be rejected.

" That is to say , the structure is a political ordering of synonym classes ; or a directed graph. "

In Fig.6.3.4.3 below, the prepositional phrase is correctly parsed, though there is no comma between "reason" and "the".

" For this reason the program has been written to accomodate batched requests , and will handle as many as 225 at one pass."

The reason is that the determiner can not be modified from the left side, so the relation between "reason" and "the" is cut off at that point. And also in this case the comma just before "and" is not necessarily needed, because the phrases on both sides of the conjunction are different from noun phrases, and they are connected correctly by the rule "VE·C1·VE ⟶ VE" or "VE·C3·VE ⟶ VE".

6.4 Pre-edit and post-edit

In the previous section, several types of erroneous samples were given. These samples are in most cases very difficult to be correctly processed only by machine. On the contrary, such cases are comparatively easy for man to recognize their real grammatical structure even from the

deformed appearences. Therefore, if slight transformations which are very easy for man are applied to the input sentences, it is apparent that even the mechanical translation can be of use. Then, what kind of transformations are effective? It may be a very good idea to compel the original sentences to be written in such styles as can be easily translated by machine, but this restriction is too severe to the writer. Therefore, the more generous treatment must be suggested in which English sentences written in an usual way are amended without any disturbance of the original word order by any person who knows only English, when he reads them only once in one direction. Several important cases are explained in the following.

(1) On the case of verbs.

The most reliable method to determine the function of verbs is to put a special sign on the verbs before the sentence is fed into the computer, though it may be somewhat tedious if there are many verbs in one sentence. But it does not necessarily mean that such signs are attached to all verbs, because such words as functional verbs (be, have, do, may, will, etc.), or ing-form verbs and past-form verbs which are used as adjectival words, passives, progressives or perfect tenses etc., can be easily recognized as verbs from the limitted context. For example, as shown below,

" The search *proceeds to *test the next document index. "

a special sign (*) is added to the words which work as verbs in the sentence. The process of translation is the same as the ordinary one mentioned in the early section, but as for the word with the sign * its part of speech is determined as verb in disregard of the order of part of speech registered in the dictionary. In the next example all verbs

the structure complex. Therefore several editting rules must be introduced as to the use of comma and hyphen.

(a) When several phrases of the same kind are arranged, the form must be as "A, B,...., and X".

(b) If a conjunction connects sentences, a comma must be placed just before the conjunction (S, and S).

(c) If the end of phrase can not be determined clearly from the syntactic information, a comma must be inserted there.

(d) It is desirable to eliminate a hyphen in the next sentence.
" band width compression system for both black-and-white and colour television " or " with the present state-of-the-art"

(e) Such a hyphen as shown below need not be eliminated.
" into a three-word representation ", or " in the large-scale study"

(f) It is advisable to insert a hyphen in such cases as shown below.
" information-processing machine " or "a punched-card and delayed-access memory and batched-process "

There are several other cases which need some pre-edition, such as treatment of parenthesis or quotation mark, idiomatic expressions which seem to be difficult for the machine, etc. These forms differ from case to case, so they can not be definitely described, but only shown by examples. Several examples of pre-edition are shown in the next page, where V mark means deletion, ∧ means insertion of comma, and (I)^{1} means that I is substituted by 1.

Table 6.4.1 Example of pre-edition

Various types of system blueprints have been discussed, and some of the questions *raised have generated considerable heat-notably the problem of financing and the related question of government vs. private control. Another, somewhat more technical, problem is the question of geographic distribution of centers and of centralization vs. decentralization. This situation is summarized in table I, where an example is given first of geographically decentralized units, next of a discipline-oriented plan where only a single center *stores the main materials *pertaining to a given subject area, and finally of a hybrid setup.

The centralization issue is a difficult one to *deal with. On the one hand, it has often been claimed that centrally managed facilities could lead to more rational operations, particularly (since) the vexing questions of compatibility between centers, and of cooperation and interaction among a variety of semi-independent units would not then *arise. On the other hand, at least one study *exists which *uses a mathematical model to *show that centralized information facilities could not operate satisfactorily, because of the great bulk of information which must then be handled in one central location [10]. (from " Proc. of IEEE, vol. 54, no. 12, p 1666-r, 1966")

6.5 About the translated Japanese

In section 6.3, mainly the results of English syntax analysis were discussed, and now the translated Japanese will be discribed here. It is interesting to investigate what kind of translations on earth

were obtained, although even nonsense translations are not minded because from the very beginning of the study the semantic information is supposed not to be used. In Appendix A some of them are shown. Only one translation word is selected from an ordinary English-Japanese dictionary for each part of speech which each English word has, and registered in the word dictionary without any semantic information. Therefore, however equivocally the word is used in the real situations, only one meaning which is used comparatively frequently is given to the word, if the grammatical function is unique. This is the same concerning the preposition which changes its expression in Japanese according to the meaning of the word related to it: therefore, it can not be helped that the translated sentences are rather nonsensical when there are many equivocal words or function words. For example, the next is a translated Japanese for a sentence which was correctly analyzed in syntax.

Japanese; sono gooseino kaikyu no sono sizen sosite
kino wa sono tansaku puroguramu o gironsuru
sono setudan no naka ni kijyutuserareru darowu.

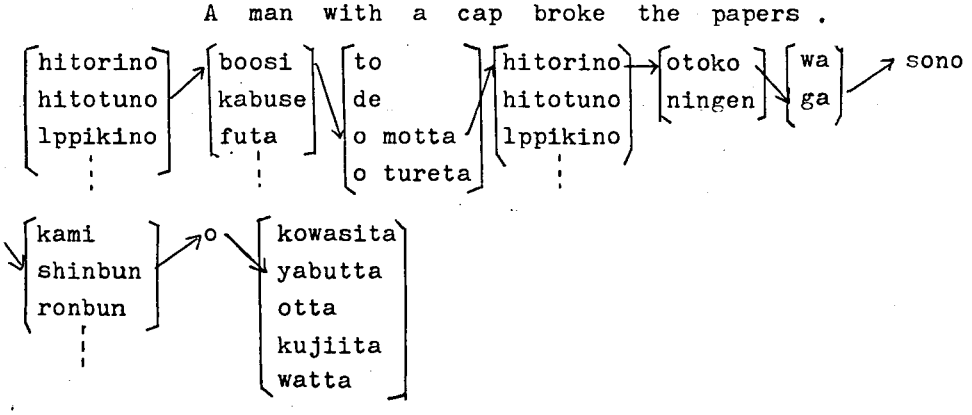
English ; The nature and function of the compaund class
will be described in the section which discuss
the search program.

It may be difficult to understand the meaning of the whole sentence from this translation, but if we substitute equivocal words with adequate ones without changing the word order in the above result, it become somewhat readable and understandable, as shown in the next page.

Japanese; fukugo kaikyu no seisitu to kino wa tansaku puroguramu o atukau syo ni kijyutusareru.

複合階級の性質と機能は探索プログラムを扱う章に記述される.

One method to improve this situation is to print out several translation words for equivocal words, and make a reader select an adequate one out of them. As for the preposition (in), such method has already been adopted in this system, and its translation is "N1(NO)", "N1" or "NO" is selected by the reader according to the semantical function of the prepositional phrase, that is, adverbial or adjectival. As regards the ordinary words, to print out several possible translation words is rather complex to edit afterwards, and much more for prepositions, although in the present state there is no other good method than this post-selection by man to treat semantic information as simple as possible. A simple sample is shown below. It is to be desided that this post-edition is to be carried out without referencing original English



sentences, but unlike Russian-English translation, it is very difficult to do so only by reading stiff Japanese in English-Japanese translation.

In the previous example (Fig.6.5.1), the translation words for "the (SONO)", "a (1NO)" and "an (1NO)" appear very frequently and they are hard to read because in ordinary Japanese such words seldom appear except in special cases. It is very easy to erase such words from the translated Japanese, namely, it is only necessary to give such words (the, a, an) a blank word instead of "SONO" and "1NO" when a mechanical dictionary is constructed.

Even in syntax-to-syntax translation, one-to-one correspondence between the source language and the target language is kept in the sense that all words except particles in Japanese have their corresponding English words, as in word-for-word translation. This comment is not trivial, for in human translation one word does not necessarily correspond to one word, but sometimes corresponds to more than two, or inversely many words may correspond to one word.

That is, schematically speaking, the same domain can be segmented into several sub-domains as in Fig.6.5.2. In other words, a set of words in both languages correspond to each other, and an English component word does not necessarily correspond to a Japanese component word. This is one of the reasons why the results of mechanical translation from English into Japanese is thought to be awkward, and also the translation itself is difficult. Furthermore, connective words which represent the relation between phrases to be connected can not be adquately determined by the syntactic information only, this also contributes to the

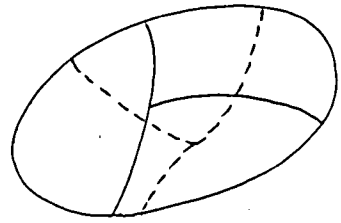


Fig. 6.5.2

awkwardness of mechanical translation. Both results are compared in the next sample, where human translation closely follows the original structure of English.

English ; In order to make change in the logic structure, including those changes which are required to create the file initially, a set of requests is provided.

Machine translation ; sono ronri koozoo no naka ni(no) henka o tukuru tameni, hajimeni fairu o soozoosuru tameni hituyootoserarereru sorera henka o fukumitutu, yookyu no 1no kumi wa yooiserareru.

Human translation ; hajimeni fairu o tukuru noni yoosuru henko mo fukumete, ronri koozoo o kaeru niwa, ichiren no situmon ga hituyoo to naru.

The above example may not be suitable for explaining the awkwardness of mechanical translation. But even if individual translation words are adequate in that sentence, there are many cases where the translation is not understandable or stiff because relational words do not reflect the relation between phrases or clauses. The suitable selection of these words is tremendously difficult. In these points there is a negative possibility that mechanical translation can not be of practical use, though syntactical analysis may be applied to other purposes, for example, a question analysis in an information retrieval system.

Chapter 7

CONCLUSION

The problem studied in this thesis was a mechanical translation of English syntax into that of Japanese using a general-purpose digital computer. The fundamental attitude is that both English and Japanese are supposed to be phrase-structure in syntax, and classification of part of speech, construction of rewriting rules, and parsing algorithm were studied, based on the belief that the analysis of the source language must depend on the character of the target language.

It is doubtful, in the strict sense, that English and Japanese have phrase-structure, but the hypothesis may be effective for the ordinary scientific papers from the practical point of view. It is obvious, however, that their structures are not context-free, and then they must be treated as context-sensitive, but to store various kinds of contexts or usage is not of practical use. Therefore, such information was taken into consideration to make formally context-free rules practically context-sensitive. That is, characteristic features of English and Japanese syntax should be reflected on the rewriting rules in some aspects: (1) in determining parts of speech or symbols for English words and phrases, (2) in giving hierarchy to each rule, (3) in parsing the structure from right to left. By these rules, analysis of the source language and synthesis of the target language can be easily performed. As regards ambiguous structures, one of the possible cases is obtained. In this case, it is a very important problem whether it is possible to restrict the number of rules and also to unify the form of each rule, that is, to limit the length n in $\alpha_1 \cdot \alpha_2 \cdots \alpha_n \rightarrow \beta$. It may be said

that in English n can be 2, but in the case of English-Japanese translation, there appear several cases which require n to be four. Although it depends on the performance of the computer, to process the rules of different length by the same algorithm needs extra time. Therefore, the length of rules must be made as short as possible. In this paper the length was restricted to only two or three, ignoreing some cases which need four. This restriction, however, was comparatively satisfactory. But if the computer speed becomes faster, it is not necessary to limit the length.

As for the hierarchy of rewriting rules, they were divided into two major classes according to their grammatical function, and each class was further divided into two sub-classes according to the length of rules. Though there are only four classes, they are equivalent to at least eight classes because they are repeatedly used by the parsing algorithm. To tell the truth, it is desirable to set order to every rule, but it may be rather wasteful. One of the problems is to know whether there is a systematic method to classify rewriting rules into several hierarchies. In natural languages, however, unlike in the formal one, the hierarchy of rules can not be determined absolutely, but it is only relatively determined, investigating various real samples.

In this paper, the length of rules was limited to two or three, and the hierarchies were also limited formally to four. The reason is that the object of this thesis was to investigate what kind of syntactic translation could be obtained in the most severe condition, therefore if this condition is proved to be too severe, then it is gradually to be loosened. It may be concluded that it is possible in outline to analyze English syntactic structures and synthesize corresponding Japanese structures by the method and dictionaries described in this

paper, if several minor points are excluded. As for parts of speech, however, they need to be classified into minor classes and their function must be more definitely determined in a given sentence. And further, there were several immoderate cases which could not be treated by such stiff context-free phrase-structure rules, and so a separate treatment was sometimes necessary.

The characteristic points of this method are summed up as follows:

(1) This method is oriented to English-Japanese translation.

(2) Context-free rules are used according to their hierarchy.

(3) Right to left scanning is adopted in parsing, unlike in an ordinary method.

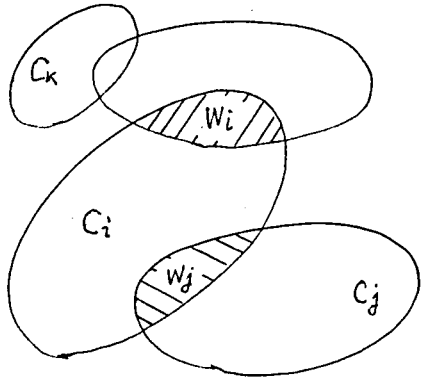
(4) Program is independent of grammar.

(5) This is syntax-to-syntax translation.

Now, the reason why mechanical translation is difficult is considered below from the general aspect, though one reason was described in section 6.5 from the syntactic point of view.

In the present state, only the information which the machine can utilize is concerned with the functions in syntactic structure. However, such information is not necessarily needed to understand the sentence, but inversely it is determined by the contents of the sentence. Therefore, to understand the contents of the sentence is most necessary, but we can not give enough informations to the computer. Though various kinds of information can be given partially, they are too incomplete for the computer to be able to associate a certain event with other ones (association mechanism is the most important function of human brain activity). Syntactic information treat surface relation of subjects, but it does not give mutual deep relation, as if it were a frame of cross word puzzle which provides mutual positions of words, but

does not give the words themselves. When we perceive or think of something, such recognition may consist of several units of minimum meaning which may be called semantic-element. Namely, concept of a certain word (Wi) is recognized as supperposed portion of several unit concepts (Ci,Cj,Ck..). Therefore, Wi and Wj which have the common concept (C_i) will give us a impression that they are mutualy related, and when one of them appears, the other one is associated with it and both of them come to co-exist in our brain, because they are connected by the rings of fundamental concepts. Then, is it possible to give the fundamental concept to the computer? The answer is rather negative, because there exist no fixed concepts, but only relative concept exist which depends on the subject, just as no one measures the distance between Earth and Venus down to the unit millimeter. Therefore, it is practically impossible to provide all fundamental concepts in advance taking all situations into consideration. Then, taking it the other way round, the concept (C_i) can be constructed by the set of words (Wi, Wj...). That is, a set of words which have the similar aspect when looked from a certain angle is thought to be a concept. This concept can not be expressed explicitly, but can be used through the words belonging to the same group. To do this, however, association algorithm which constructs the word group of similar kind from the given verbal data must be developed. Of course some information other than samples of usage must be given directly, but the computer can gradually get several groups of words. In this case, syntactic analysis



is useful in investigating the positional relation among some words.

The syntactic analysis method, which is studied in this paper, may aid several language processing systems, such as automatic abstracting, contents analysis, sentence generation, and also man-machine communication especially in information retrieval.

ACKNOWLEDGEMENTS

This doctoral thesis has been studied under the guidance of Professor Toshiyuki Sakai. The author thanks him from the bottom of his heart for his instructive advice, inspiring encouragement, and fatherly affection during the course of this investigation. He also thanks Mr. Makoto Nagao for his suggestions and helpful discussions about many problems in language processing.

He is also very much indebted to the staff of Professor T. Sakai's laboratory, especially Assistant Professor S. Doshita and H. Nishio for valuable discussions at every opportunity.

He is much obliged to Professor Yukinobu Oda of Doshisha Women's College for correcting ungrammatical expressions. The author is afraid that too many mistakes would make his eye blind.

Last, but not least, the author must thank to Miss Kimie Miyake for her self-sacrificing efforts in typing this manuscript.

BIBLIOGRAPHY

Chapter 1

(1) Bar-Hillel, Y. "The Present Status of Automatic Translation of Languages", in F. Alt (ed.) Advances in Computers I, Academic Press, 1960.

(2) Hjelmslev, L. "Prolegomena to a theory of Language", Indiana University publications in anthropology and linguistics, Memoir 7 of the International Journal of American Linguistics, Supplement to Vol. 19, No.1, 1953.

(3) Hume, D. "A treatise of human nature", Reprinted from the original edition. Edited by Selby-Bigge, L.A. Oxford, Clarendon Press, 1888.

(4) Miller, G.A. "Language and Communication", McGraw-Hill paperbacks, 1963.

(5) Nida, E.A. "Toward a Science of Translating", Leiden E.J.BRILL, 1964.

Chapter 2

(6) Booth, A.D. (ed.), "Machine Translation of Languages", the MIT Press and John Wiley & Sons, 1957.

(7) Booth, A.D. (ed.), "Machine Translation", North Holland, 1967.

(8) Brower, R.A. (ed.) "On Translation", Harvard Univ. Press, 1959.

(9) Chomsky, N. "Syntactic Structures", Mouton and Co., The Hague, 1957.

(10) Chomsky, N. "Aspects of the theory of syntax", the MIT Press, 1965.

(11) Delavenay, E. "La Machine a Traduire", collection QUE SAIS-JE ? N 834.

(12) Gross, M. "On the Equivalence of Methods of Language Used in the Fields of Mechanical Translations and Information Retrieval", presented at the NATO Advanced Study Institute on Automatic Translation of Languages, 1962.

(13) Hays, D.G. "On the value of Dependency connection", 1961 International Conference on Machine Translation of Language and

Applied Language Analysis (ICMTLALA), Her Majesty's Stationery Office, London, 1962.

(14) Kuno, S. and Oettinger, A.G. "Multiple-Path Syntactic Analyzer", Proce. IFIP 62, North Holland, 1963.

(15) Oettinger, A.G. "Automatic Language Translation", Harvard Univ. Press, 1960.

(16) Panob, D.V. "Automatic Translation", USSR Science Academy, 1956.

(17) Rhodes, P. "A new model of syntactic description", 1961 ICMTLALA (ref. (13)), Her Majesty's Stationery office, London, 1962.

(18) Simon, R.F. "Answering English Questions by Computer: A Survey", Comm. of ACM, Vol. 8, No.1, 1965.

(19) Sugita, S. "Automatic Language Processing", mineograph, January, 1966.

Chapter 3

(20) Francis, W.N. "The Structure of American English", The Ronald Press Co., N.Y., 1958.

(21) Fries, C.C. "The Structure of English", Longmans, Green and Company, London, 1961.

(22) Harvard University "Language Date Processing", Harvard summar school paper, 1964.

(23) Hornby, A.S. "A Guide to Patterns and Usage in English", Oxford Univ. Press, 1957.

(24) Kleinjans, E. "A Descriptive-Comparative Study Prediction Interference for Japanese in Learning English Noun-Head Modification Patterns", Michigan, 1958.

(25) Otuka, T. (ed.) "Sanseido's Dictionary of English Grammar", Sanseido's Japan, 1963.

(26) Roberts, P. "Patterns of English", Harcourt, Brace and Company, 1961.

(27) Strang, B.M.H. "Modern English Structure", Edward Arnold Publishers, London,1962.

(28) Sakai, T. "Models and Strategies for MT", Preprints for seminar on Mechanical Translation, April, 1964.

(29) Sakai, T., Nagao, M. and Sugita, S., "A method of English-Japanese Machine Translation", Record of the 1963 Conventions of the Information Processing Society of Japan, December, 1963.

Chapter 4

(30) Kiyono, T., Mitsumori, S. and Miyamoto, M., "Machine Translation", The Journal of the Institute of Electrical Communication Engineers of Japan, Vol. 46, No.11, November, 1963.

(31) Simmon, R.F. and Klein, S., "A Computational Approach to Grammatical Coding of English words", Journal of ACM, Vol. 10, No.6, July, 1963.

(32) Stolz, W.S., et al., "A Stochastic Approach to the Grammatical Coding of English", Comm. of ACM, Vol. 8, No.6, June, 1965.

(33) Univ. of Washington, "Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian into English", Report no. RADC-TN-58-321, ASTIA document no. AD-148992, 1958.

Chapter 5

(34) Foster, D., "Automatic Sentence Kernelization", Mathematic Linguistic Seminar Papers, Vol. 10, Harvard Univ., 1964.

(35) Sakai, T. and Sugita, S., "Mechanical Translation of English into Japanese", The Journal of the Institute of Electrical Communication Engineers of Japan (J of IECEJ), Vol. 49, No.2, pp46-53, 1966.

(36) Sakai, T. and Sugita, S., "English-Japanese Translation by Digital Computer", Technical Report of the Proffessional Group on Automation and Automatic Control of the Institute of Electrical Communication Engineering of Japan, February, 1966.

(37) Simmon, R.F., et al., "Analyzing English Syntax with a Pattern Learning Parser", Comm. of ACM, Vol.8, No.11, Nov., 1965.

(38) Tadenuma, R., "English-Japanese Machine Translation (I)", Researches of the Electrotechnical Laboratory, No.624, December, 1961.

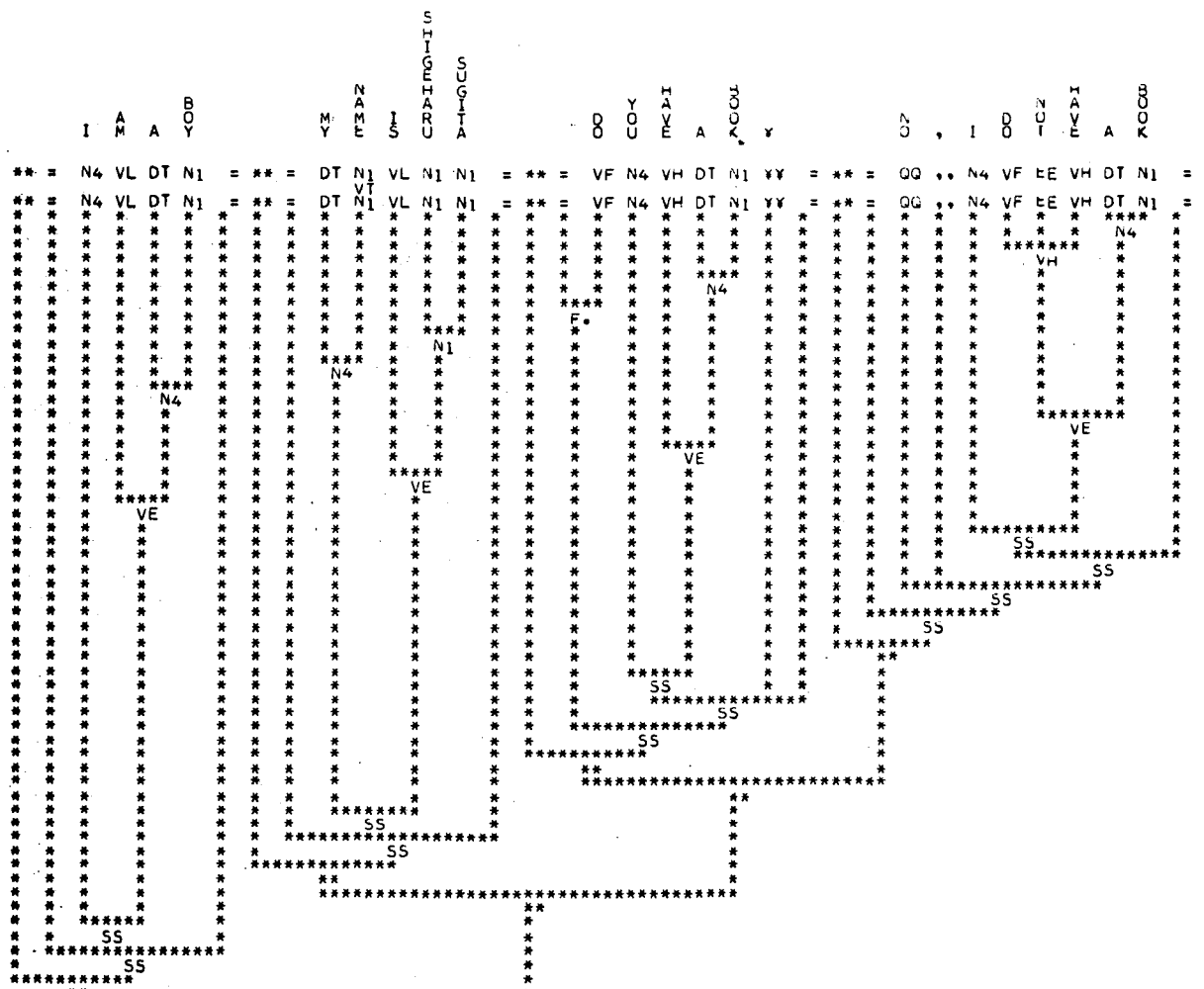
(39) Tadenuma, R. and Igarashi, J., "English-Japanese Machine Translation (II)", Researches of the Electrotechnical Laboratory, No.631, November, 1962.

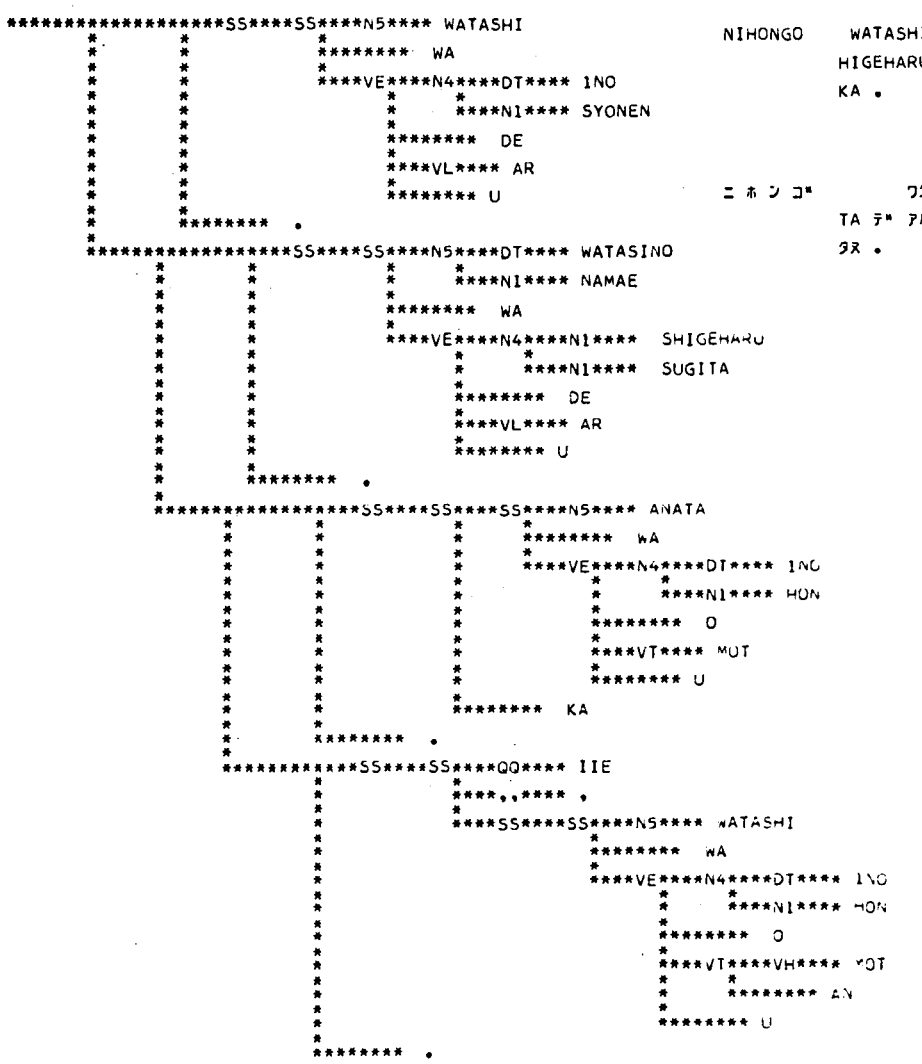
(40) Tamachi, T., "Process of Automatic Translation", Technical Report of the Proffessional Group on Automation and Automatic Control of IECEJ, June, 1959.

Chapter 7

(41) Sakai, T. and Sugita, S., "Information Value of sentences written in Natural Language", Record of the 1967 Joint Convention of the IECEJ and others, May, 1967.

ENGLISH I AM A BOY. MY NAME IS SHIGEHARU SUGITA. DO YOU HAVE A BOOK?. NO, I DO NOT HAVE A BOOK.

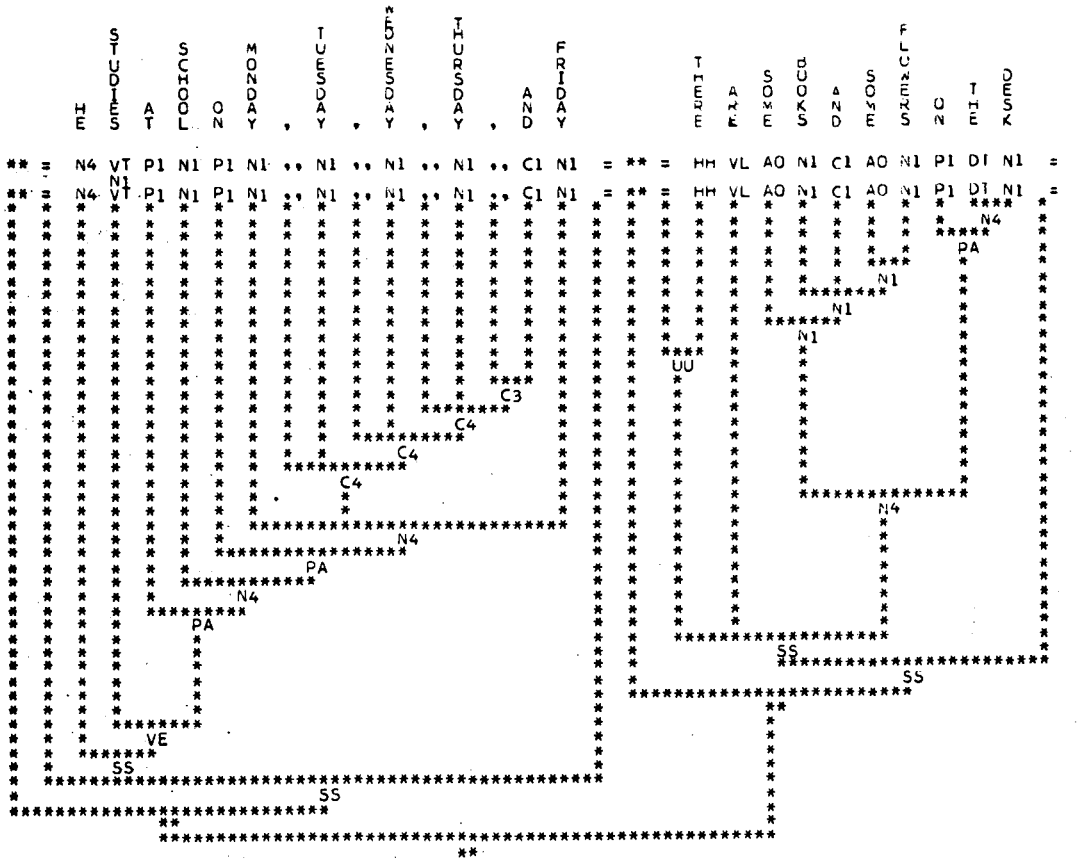


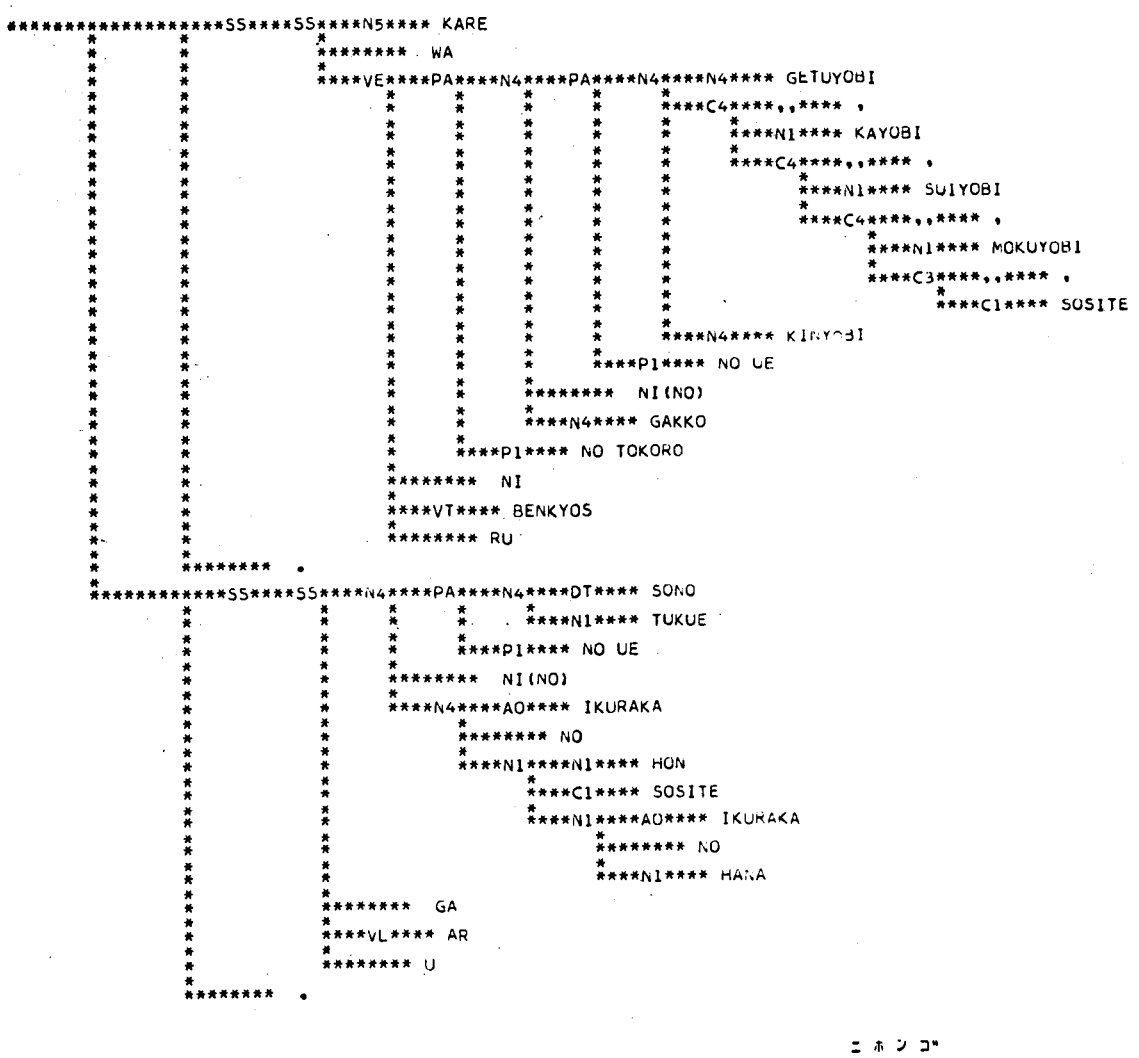


NIHONGO WATASHI WA 1NO SYONEN DE ARU . WATASINO NAMAE WA SHIGEHARU SUGITA DE ARU . ANATA WA 1NO HON O MOTU KA . IIE , WATASHI WA 1NO HON O MOTANU .

ニホンゴ ﾜﾀｼ ﾜ 1 ﾉ ｼｮｳﾈﾝ ﾃﾞ ｱﾙ . ﾜﾀｼﾉ ﾅﾏｴ ﾜ SHIGEHARU SUGITA ﾃﾞ ｱﾙ . ｱﾅﾀ ﾜ 1 ﾉ ﾎﾝ ｵ ﾓﾂ ｶ . ｲｲｴ , ﾜﾀｼ ﾜ 1 ﾉ ﾎﾝ ｵ ﾓﾀﾅｲ .

ENGLISH HE STUDIES AT SCHOOL ON MONDAY,TUESDAY,WEDNESDAY,-
THURSDAY,AND FRIDAY. THERE ARE SOME BOOKS AND SOME FL-
OWERS ON THE DESK.

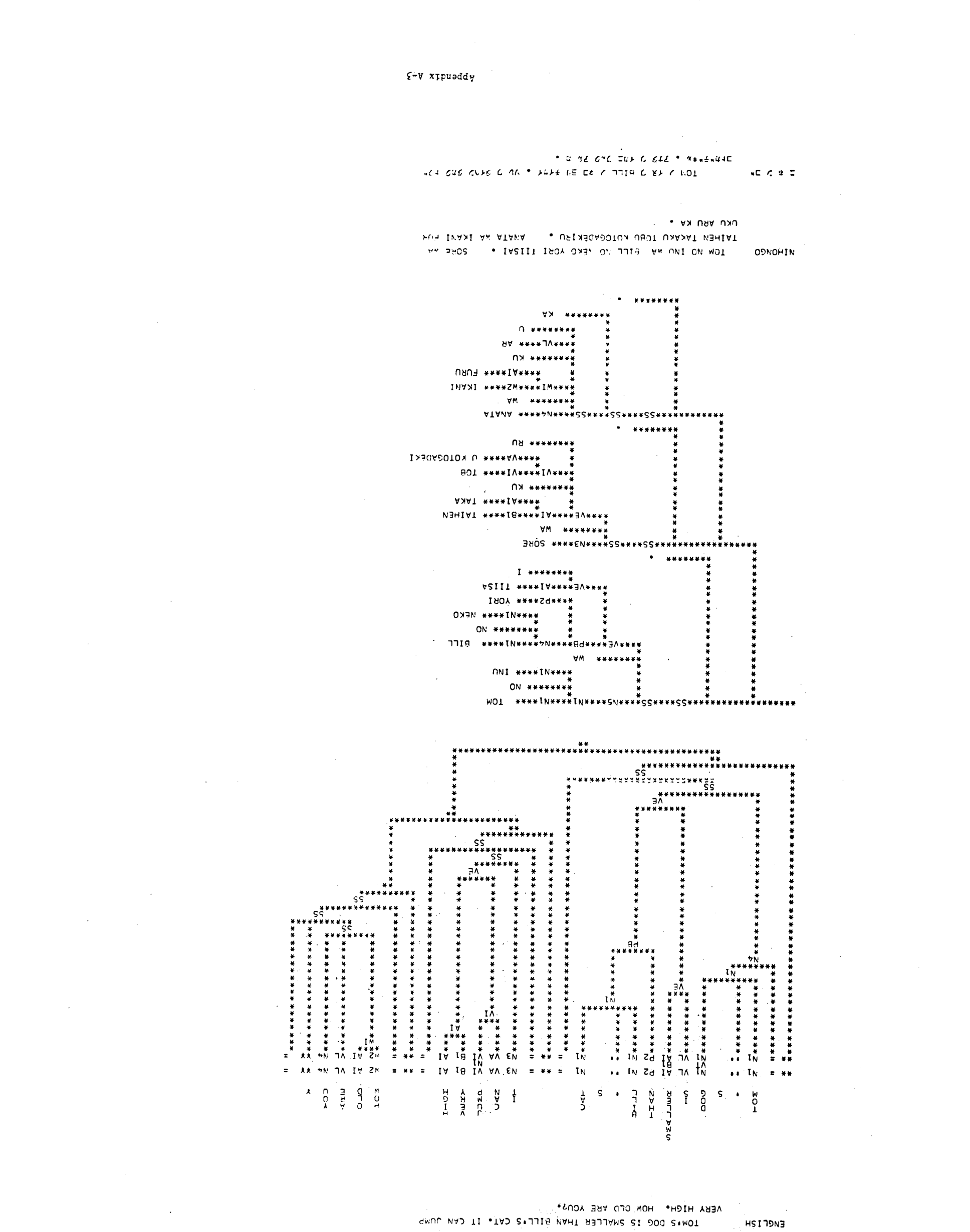




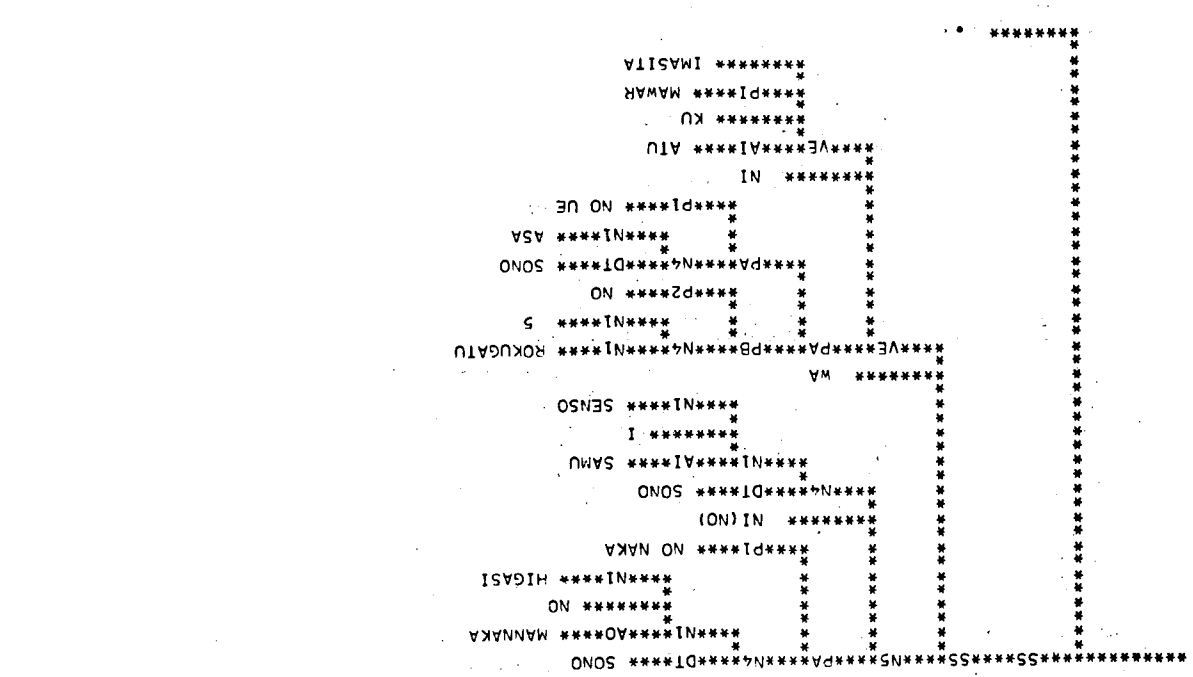
NIHONGO KARE WA GETUYOBI , KAYOBI , SUIYOBI , MOKUYOBI , SOSII
E KINYOBI NO UE NI(NO) GAKKO NO TOKORO NI BENKYOSURU .
SONO TUKUE NO UE NI(NO) IKURAKANO HON SOSITE IKURAKA
NO HANA GA ARU .

ニホンゴ

カレ ワ ｹﾞﾂﾖｳﾋﾞ , ｶﾖｳﾋﾞ , ｽｲﾖｳﾋﾞ , ﾓｸﾖｳﾋﾞ , ｿｼﾃ ｷﾝﾖｳﾋﾞ ノ
ｳｴ ﾆ (ﾉ) ｶﾞｯｺｳ ﾉ ﾄｺﾛ ﾆ ﾍﾝｷｮｳｽﾙ . ｿﾉ ﾂｸｴ ﾉ ｳｴ ﾆ (ﾉ) ｲ
ｸﾗｶﾉ ﾎﾝ ｿｼﾃ ｲｸﾗｶﾉ ﾊﾅ ｶﾞ ｱﾙ .



ENGLISH TOM'S DOG IS SMALLER THAN BILL'S CAT. IT CAN JUMP VERY HIGH. HOW OLD ARE YOU?.

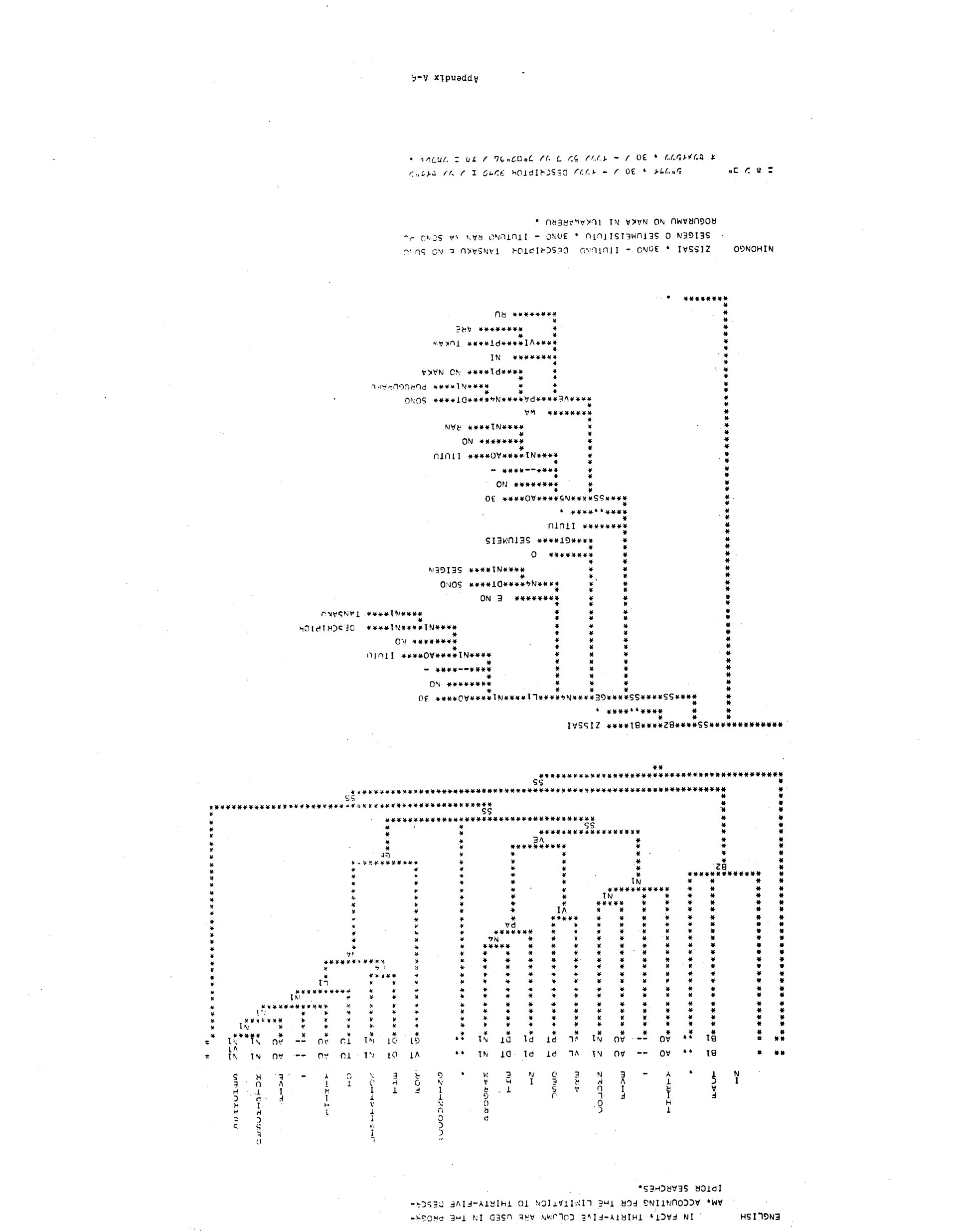


ENGLISH THE COLD WAR IN THE MIDDLE EAST TURNED HOT ON THE MORNING OF JUNE 5.

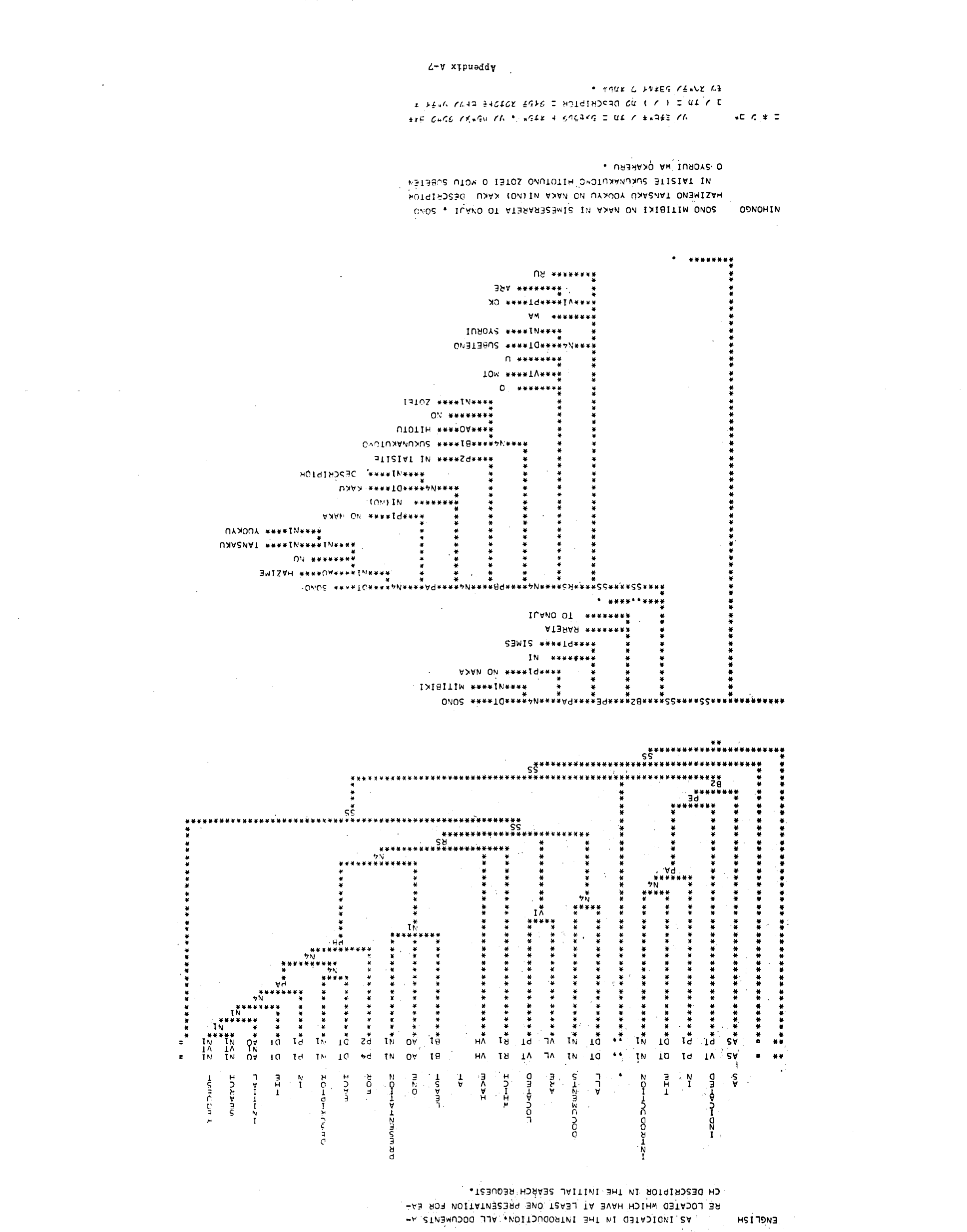
NIHONGO SONO MAINICHI NO ASA NI(NO) SONO SAMUI SENSO WA ROKUGATU 5 NO SONO ASA NO UE NI ATUKU MAWARIMASITA.

NIHONGO SONO OKURIMONO SOSIKI NO NAKA NI ATUKAWARERU KOTOGADEK
IRU SONO JYOHO MONDAI NO SONO OKISA WA TUZUKU SAMAZAMA
A SEIGEN NI YOTTE SIMESERARERU .

ENGLISH THE SIZE OF THE INFORMATION PROBLEMS THAT CAN BE
DEALT WITH IN THE PRESENT SYSTEM IS INDICATED BY VARIO-
US LIMIT WHICH FOLLOWS.



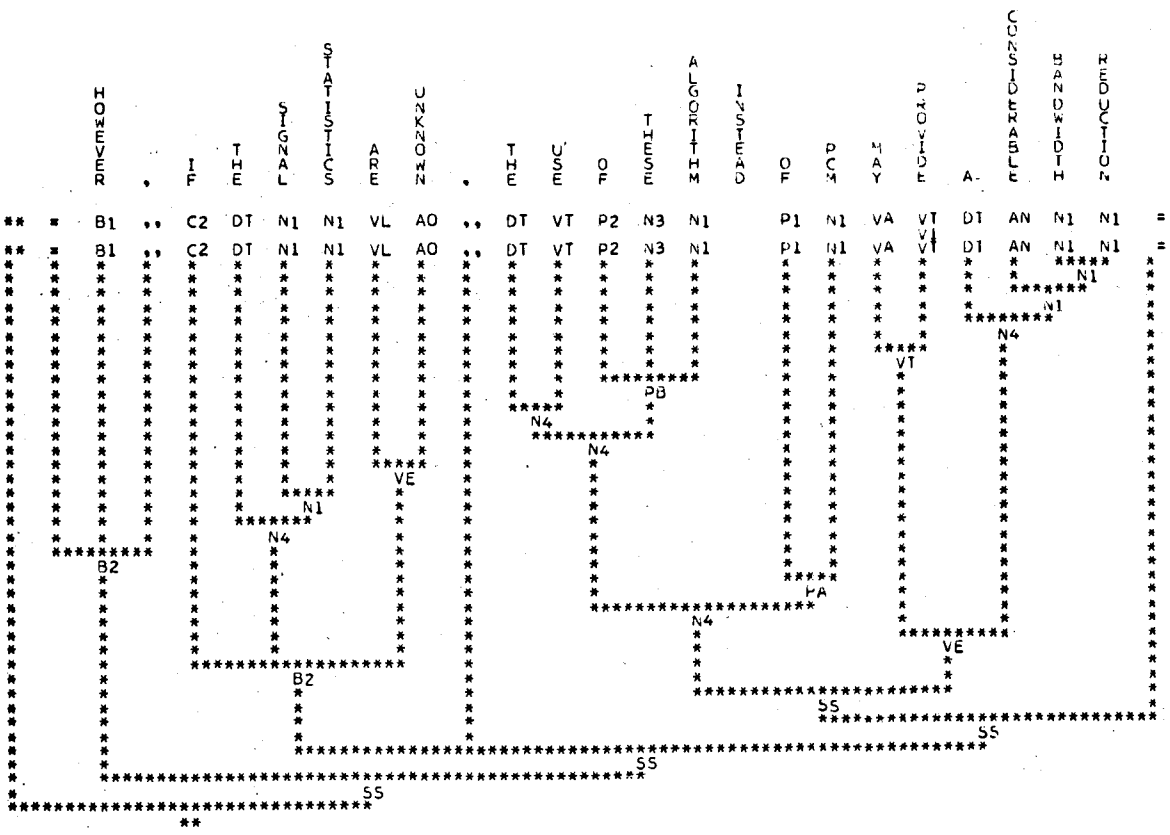
ENGLISH IN FACT, THIRTY-FIVE COLUMN ARE USED IN THE PROGRAM, ACCOUNTING FOR THE LIMITATION TO THIRTY-FIVE DESCRIPTOR SEARCHES.



NIHONGO SONO MITIBIKI NO NAKA NI SIMESERARETA TO ONAJI • SONO HAZIMENO TANSAKU YOUKYU NO NAKA NI(NO) KAKU DESCRIPTOR NI TAISITE SUKUNAKUTOMO HITOTUNO ZOTEI O MOTU SUBETEN O SYORUI WA OKAHERU •

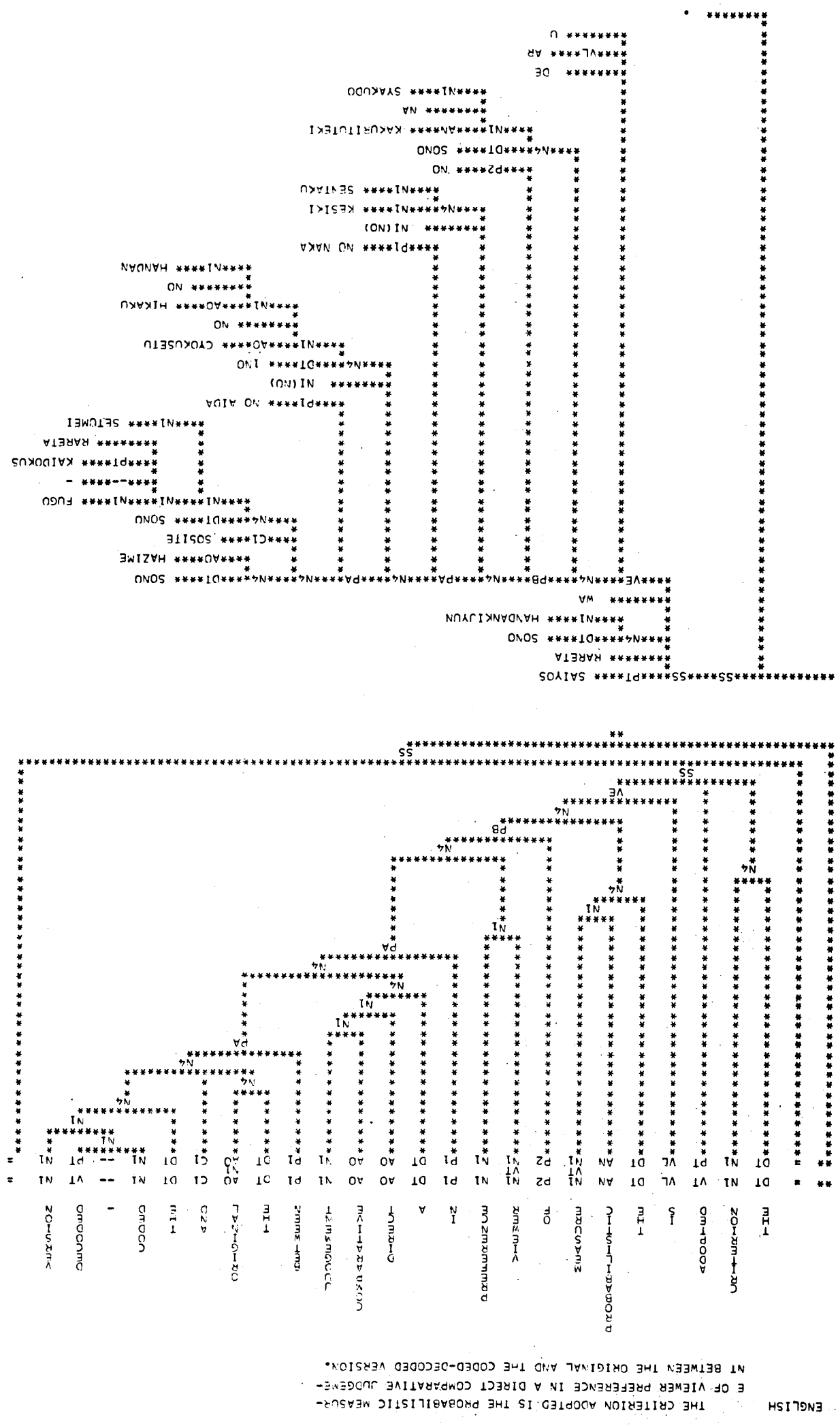
ENGLISH AS INDICATED IN THE INTRODUCTION, ALL DOCUMENTS ARE LOCATED WHICH HAVE AT LEAST ONE PRESENTATION FOR EACH DESCRIPTOR IN THE INITIAL SEARCH REQUEST.

ENGLISH HOWEVER, IF THE SIGNAL STATISTICS ARE UNKNOWN, THE USE OF THESE ALGORITHM INSTEAD OF PCM MAY PROVIDE A CONSIDERABLE BANDWIDTH REDUCTION.



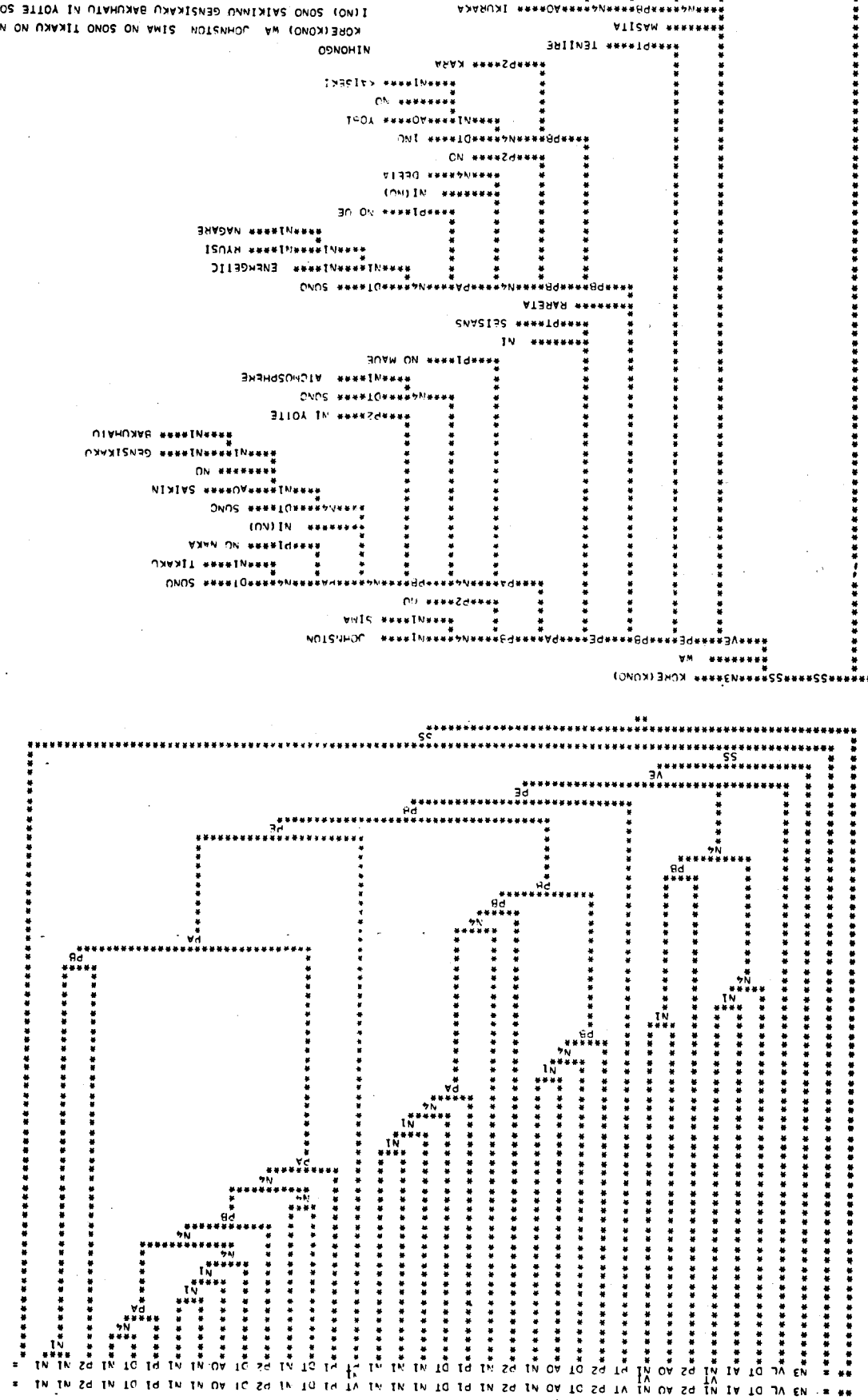
NIHONGO SIKASINAGARA , SONO SINGO TOKEI GA MITI DE ARU NARA , IT NO KAWARI NI(NO) KORERA ARUGORIZUMU NO SONO TUKAW KOTO WA INO KAWARINA TAIIKIHABA KANGEN O YOISURU KAMOSIREN .

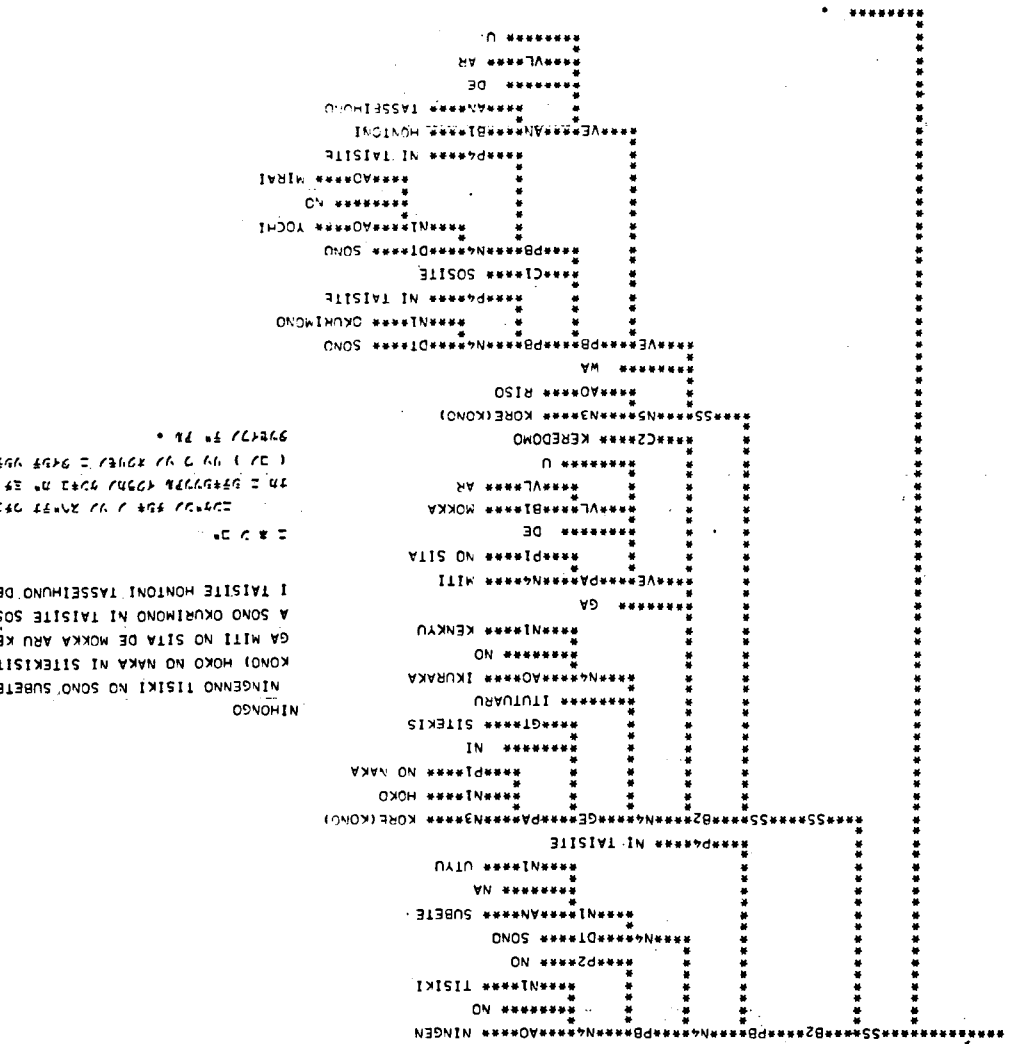
ニホンゴ シカシナガラ , ソノ シンゴウ トケイ ガ ミチ デ アル ナラ , IT ノ カワリ ニ (ノ) コレラ アルゴリズム ノ ソノ ツカウ コト ワ 1 ノ カタリナ タイキハバ カンゲン ヲ ヨイスル カモシレヌ .



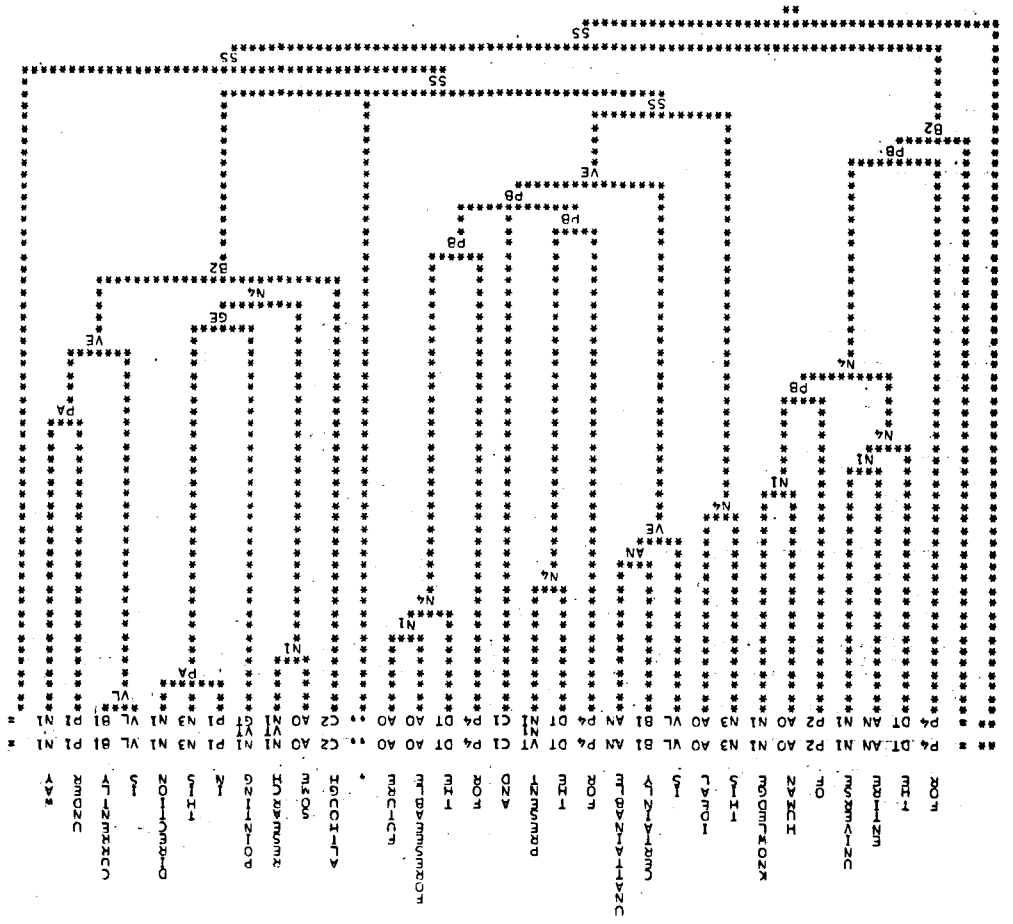
ENGLISH THE CRITERION ADOPTED IS THE PROBABILISTIC MEASURE OF VIEWER PREFERENCE IN A DIRECT COMPARATIVE JUDGEMENT BETWEEN THE ORIGINAL AND THE CODED-DECODED VERSION.

ENGLISH THIS IS A SHORT ACCOUNT OF SOME RESULTS OBTAINED FROM A PRELIMINARY ANALYSIS OF DATA ON THE ENERGETIC PARTICLE FLUX PRODUCED ABOVE THE ATMOSPHERE BY THE RECENT NUCLEAR EXPLOSION IN THE VICINITY OF JOHNSTON ISLAND.



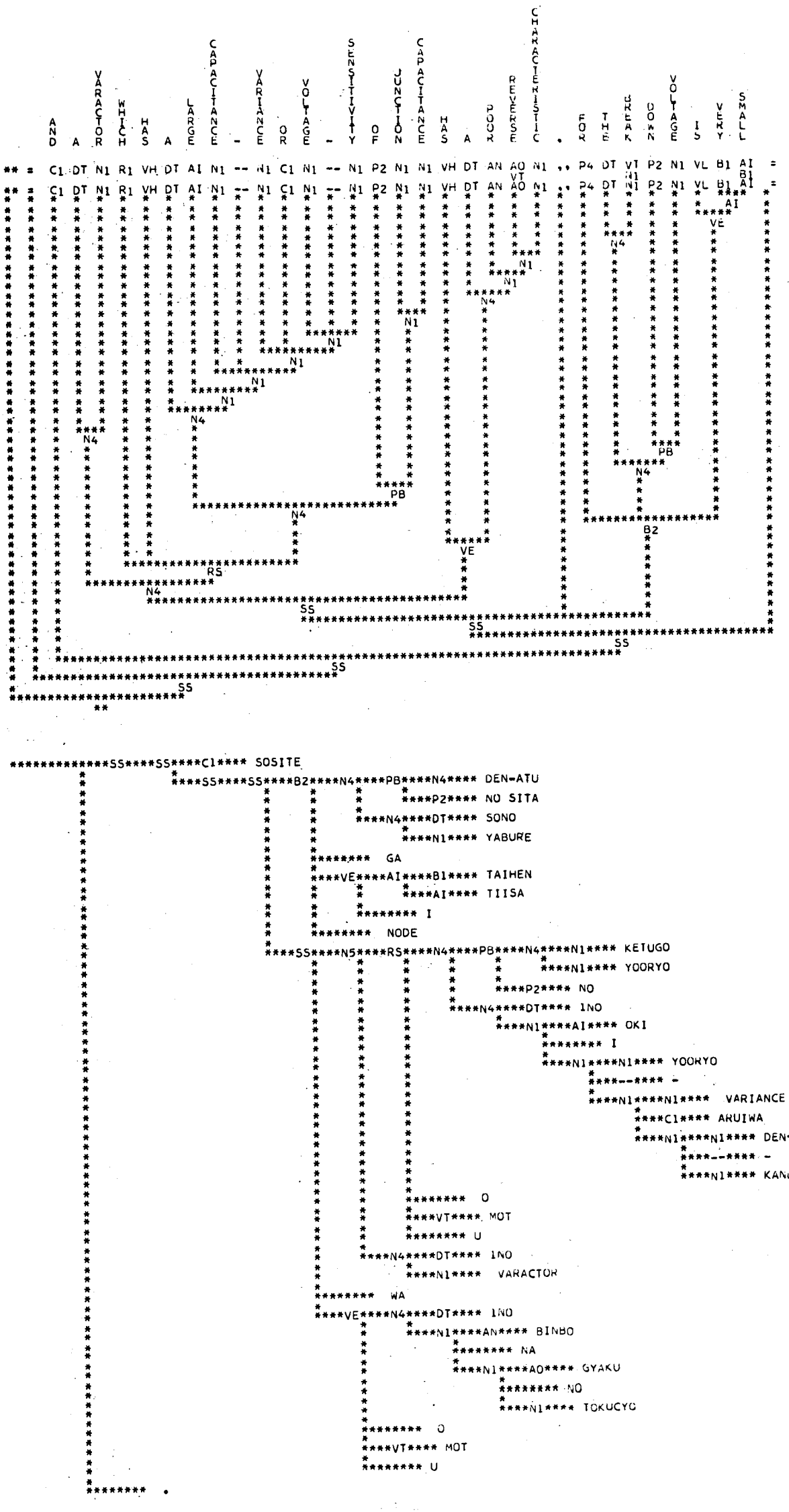


ENGLISH FOR THE ENTIRE UNIVERSE OF HUMAN KNOWLEDGE THIS I-
DEAL IS CERTAINLY UNATTAINABLE FOR THE PRESENT AND FOR
THE FORESEEABLE FUTURE. ALTHOUGH SOME RESEARCH POINTI-
NG IN THIS DIRECTION IS CURRENTLY UNDER WAY.

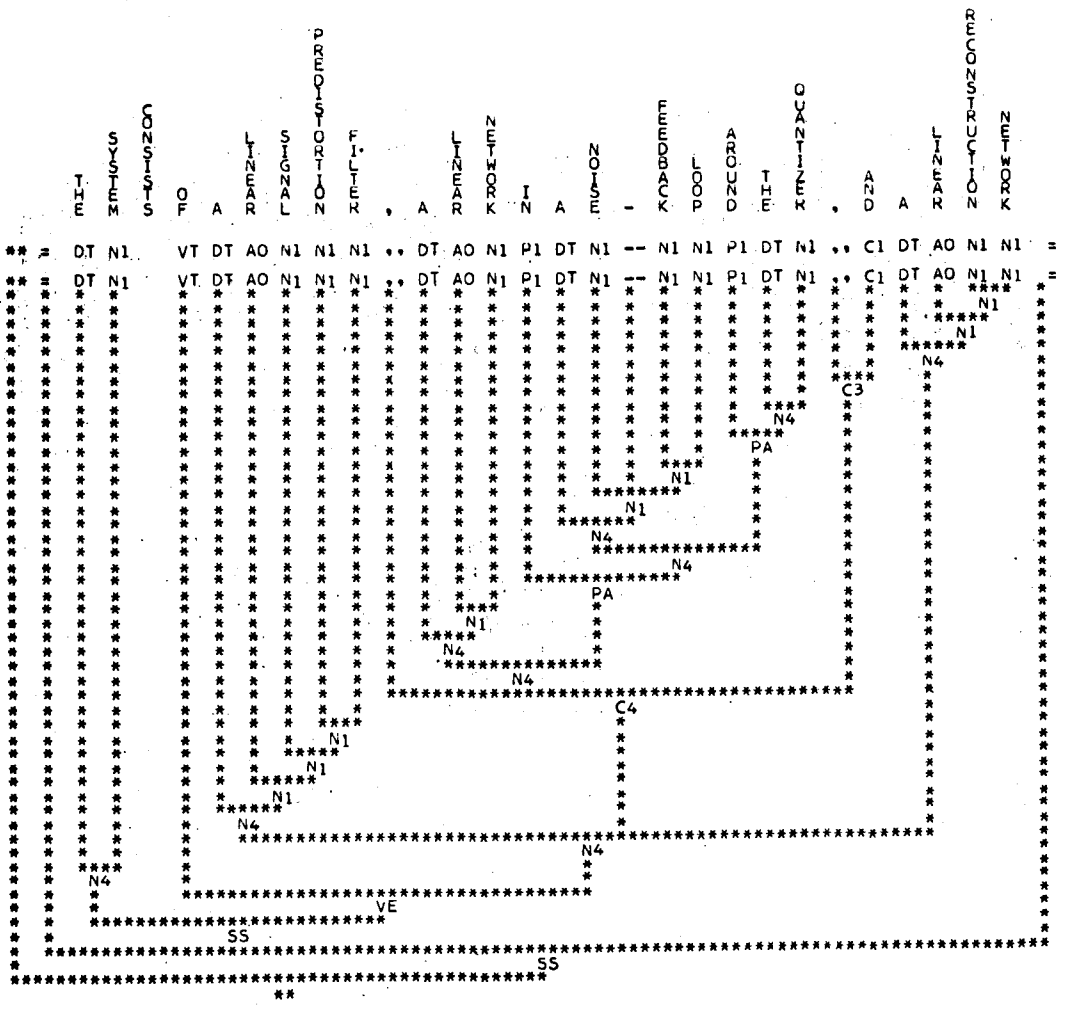


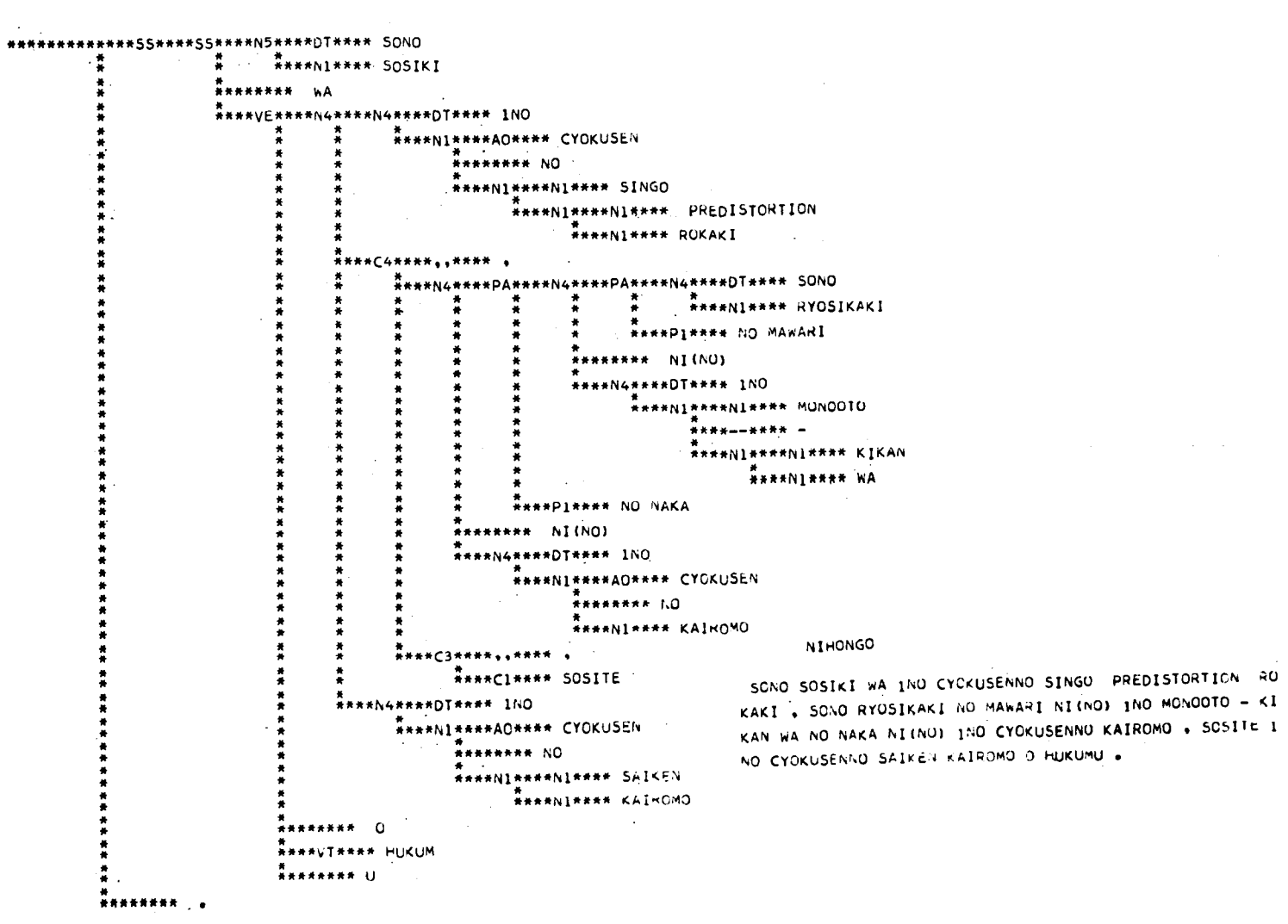
NIHONGO
NINGENNO TISIKI NO SONO SUBETENA UIYU NI TAISITE KORE(
KONO) HOKO NO NAKA NI SITEKISITUARU IKURAKANO KENKYU
GA MITI NO SITA DE MOKKA ARU KEREDOMO KORE(KONO) RISO #
A SONO OKOURIMONO NI TAISITE SOSITE SONO YOCHINO MIRAI N
I TAISITE HONTONI TASSEIHUNO DE ARU.

ENGLISH AND A VARACTOR WHICH HAS A LARGE CAPACITANCE-VARIANCE OR VOLTAGE-SENSITIVITY OF JUNCTION CAPACITANCE HAS A POOR REVERSE CHARACTERISTIC. FOR THE BREAK DOWN VOLTAGE IS VERY SMALL.



ENGLISH THE SYSTEM CONSISTS OF A LINEAR SIGNAL PREDISTORTION FILTER, A LINEAR NETWORK IN A NOISE-FEEDBACK LOOP AROUND THE QUANTIZER, AND A LINEAR RECONSTRUCTION NETWORK.





NIHONGO

APPENDIX B (Obtained from about 11,000 words)

CONNECTION TABLE OF PARTS OF SPEECH 2--SYMBOL

		FREQ.			FREQ.			FREQ.			FREQ.
NN	NN	0825	DT	NN	0759	NN	PP	0676	AJ	NN	0515
PP	DT	0449	PP	NN	0353	NN	**	0324	NN	,,	0299
DT	AJ	0270	NN	VL	0205	VL	VP	0201	VP	PP	0141
PP	AJ	0128	**	DT	0127	NN	C1	0121	PN	NN	0105
NN	))	0096	,,	C1	0095	NN	VV	0093	VV	DT	0090
C1	NN	0085	TO	VV	0083	--	NN	0078	VP	NN	0077
((	NN	0075	**	PN	0074	NN	VP	0072	NN	TO	0068
,,	NN	0067	AJ	AJ	0064	,,	DT	0063	PP	PN	0062
VV	NN	0061	C1	DT	0058	VV	PP	0057	VG	NN	0057
NN	((	0056	VL	DT	0053	**	NN	0053	PN	VL	0053
P4	DT	0052	NN	--	0052	NN	AJ	0050	VP	TO	0049
VL	AJ	0049	NN	P4	0048	VP	**	0047	VL	AD	0047
NN	RP	0047	AD	AJ	0046	TO	DT	0045	NN	VG	0045
P4	NN	0044	VG	DT	0040	TO	NN	0040	**	PP	0039
NN	AD	0039	AD	PP	0039	PP	VG	0038	AJ	--	0038
AD	,,	0037	NN	VA	0036	AJ	TO	0036	AJ	PP	0036
VA	VL	0035	RA	NN	0035	TO	VL	0034	,,	PP	0033
DT	VP	0033	**	AD	0032	PL	VP	0032	VP	P4	0031
NN	VH	0031	AD	VP	0031	**	AJ	0030	AJ	,,	0030
AD	DT	0030	,,	AD	0028	NN	DT	0028	HH	VL	0027
VV	AD	0026	C1	AJ	0026	VV	TO	0025	PN	VV	0025
,,	VG	0024	VP	AD	0024	RP	VV	0024	VH	PL	0023
DT	VG	0023	C1	AD	0023	AD	VV	0022	VG	PP	0021
TH	DT	0021	DT	AD	0021	AJ	C1	0021	AD	AD	0021
**	P4	0020	VP	,,	0019	VA	VV	0019	NN	AS	0019

CONNECTION TABLE OF PARTS OF SPEECH 2--SYMBOL

		FREQ.			FREQ.			FREQ.			FREQ.
..	PN	0018	VP	TH	0018	DT	RA	0018	VV	TH	0017
))	**	0017	VV	AJ	0016	PP	RP	0016	VP	AJ	0015
VL	EE	0015	TO	AJ	0015	PN	AJ	0015	P4	PN	0015
C1	VV	0015	AD	TO	0015	VH	DT	0014	TH	PN	0014
RP	VL	0014	PU	VV	0014	NN	TH	0014	NN	PN	0014
NN	PL	0014	AS	DT	0014	AS	AJ	0014	AJ	**	0014
..	P4	0013	VL	TO	0013	**	HH	0013	NN	PU	0013
AS	NN	0013	AJ	AS	0013	AD	NN	0013	..	VP	0012
..	AJ	0012	VG	AJ	0012	TH	VL	0012	PP	VP	0012
PN	VA	0012	P4	AJ	0012	))	..	0012	))	VV	0012
DT	VV	0012	AS	VP	0012	AJ	P4	0012	..	RP	0011
VV	**	0011	VV	PN	0011	PP	AD	0011	P4	VG	0011
--	AJ	0011	C2	DT	0011	AD	RA	0011	..	VV	0010
..	AS	0010	VV	..	0010	VL	..	0010	VL	NN	0010
VG	TO	0010	TH	NN	0010	RP	DT	0010	RA	PP	0010
RA	AJ	0010	PU	VL	0010	))	NN	0010	AD	**	0010
AD	PN	0010	VP	AS	0009	VG	PN	0009	**	((	0009
**	C2	0009	EE	AJ	0009	DT	DT	0009	..	((	0008
..	VL	0008	WW	DT	0008	VL	RA	0008	VH	VP	0008
VA	AD	0008	RP	VH	0008	PN	PU	0008	))	DT	0008
C1	VG	0008	AJ	VP	0008	AD	AS	0008	VP	DT	0007
VP	C1	0007	PP	PP	0007	PN	VH	0007	))	VL	0007
))	C1	0007	EE	VV	0007	C1	PN	0007	AS	AD	0007
AJ	VL	0007	AJ	VG	0007	((	DT	0006	..	WW	0006
..	TH	0006	..	C2	0006	VV	((	0006	VV	C1	0006

CONNECTION TABLE OF PARTS OF SPEECH 4--SYMBOL

				FREQ.					FREQ.					FREQ.					FREQ.
					NN	PP	DT	NN	0161	DT	NN	PP	NN	0099	PP	DT	NN	NN	0094
NN	PP	NN	NN	0093	DT	NN	PP	DT	0089	PP	DT	AJ	NN	0080	PP	DT	NN	PP	0077
DT	AJ	NN	PP	0066	NN	PP	DT	AJ	0062	NN	**	DT	NN	0061	NN	PP	AJ	NN	0059
AJ	NN	PP	NN	0049	DT	AJ	NN	NN	0048	NN	NN	PP	DT	0043	VP	PP	DT	NN	0039
NN	PP	NN	PP	0038	DT	NN	NN	PP	0038	NN	NN	**	DT	0037	NN	NN	PP	NN	0037
AJ	NN	PP	DT	0033	NN	((	NN	))	0031	NN	NN	,,	NN	0031	NN	NN	VL	VP	0031
PP	NN	PP	DT	0030	VV	DT	NN	PP	0029	DT	NN	NN	NN	0029	NN	NN	NN	**	0028
DT	NN	NN	**	0027	**	DT	NN	PP	0026	NN	,,	NN	,,	0026	NN	PP	NN	**	0026
NN	NN	NN	NN	0026	PP	NN	NN	NN	0025	PP	AJ	NN	NN	0025	NN	,,	C1	DT	0025
NN	PP	PN	NN	0025	**	DT	AJ	NN	0024	PP	NN	NN	**	0024	NN	VL	VP	PP	0024
NN	**	DT	AJ	0024	NN	NN	**	PN	0024	NN	C1	NN	NN	0024	AJ	--	NN	NN	0024
DT	NN	PP	AJ	0023	AJ	NN	NN	**	0023	VL	VP	PP	DT	0022	PP	NN	NN	,,	0022
NN	,,	DT	NN	0022	NN	**	PN	NN	0022	AJ	NN	VL	VP	0022	VV	PP	DT	NN	0021
TO	VV	DT	NN	0021	NN	VP	PP	DT	0021	NN	NN	NN	PP	0021	DT	AJ	NN	**	0021
AJ	NN	,,	C1	0021	VL	VP	PP	NN	0020	**	DT	NN	NN	0020	PP	DT	NN	,,	0020
NN	NN	PP	AJ	0020	DT	NN	NN	VL	0020	DT	AJ	AJ	NN	0020	,,	C1	DT	NN	0019
NN	NN	,,	C1	0019	NN	NN	NN	,,	0019	DT	NN	NN	,,	0019	P4	DT	NN	PP	0018
NN	VL	VP	TO	0018	NN	VH	PL	VP	0018	DT	NN	VL	VP	0018	PP	NN	NN	PP	0017
PP	DT	NN	**	0017	PP	DT	NN	C1	0017	NN	VV	DT	NN	0017	NN	VL	VP	**	0017
NN	PP	NN	,,	0017	PP	DT	VP	NN	0016	NN	VA	VL	VP	0016	AJ	NN	PP	AJ	0016
AJ	NN	NN	NN	0016	AD	PP	DT	NN	0016	,,	DT	NN	PP	0015	VV	DT	AJ	NN	0015
TO	DT	NN	PP	0015	NN	VL	VP	P4	0015	NN	**	PN	VL	0015	NN	NN	C1	NN	0015
C1	DT	NN	NN	0015	AJ	PP	DT	NN	0015	,,	DT	AJ	NN	0014	VP	PP	NN	NN	0014
VL	DT	AJ	NN	0014	PP	AJ	NN	PP	0014	NN	VL	DT	NN	0014	NN	**	AD	,,	0014
NN	P4	DT	NN	0014	NN	NN	NN	VL	0014	DT	NN	C1	NN	0014	,,	NN	,,	NN	0013

CONNECTION TABLE OF PARTS OF SPEECH 4--SYMBOL

				FREQ.					FREQ.					FREQ.					FREQ.
VV	NN	PP	DT	0013	VL	DT	NN	NN	0013	PP	NN	PP	NN	0013	NN	,,	NN	NN	0013
NN	VV	PP	DT	0013	NN	NN	,,	DT	0013	NN	NN	**	PP	0013	NN	--	NN	NN	0013
DT	NN	PP	PN	0013	AJ	NN	**	DT	0013	VP	PP	DT	AJ	0012	VP	P4	DT	NN	0012
VL	VP	TO	VV	0012	VA	VL	VP	PP	0012	**	PN	VL	VP	0012	PP	VG	DT	NN	0012
PP	NN	**	DT	0012	PP	AJ	NN	VL	0012	NN	**	PP	DT	0012	NN	**	NN	NN	0012
NN	**	AJ	NN	0012	NN	PP	NN	VL	0012	NN	NN	VL	AD	0012	NN	NN	**	NN	0012
NN	--	NN	PP	0012	NN	C1	DT	NN	0012	--	NN	NN	NN	0012	DT	NN	VP	PP	0012
C1	DT	NN	PP	0012	VP	**	DT	NN	0011	VP	PP	AJ	NN	0011	VG	PP	DT	NN	0011
PP	DT	AJ	AJ	0011	NN	,,	PP	DT	0011	NN	,,	DT	AJ	0011	NN	VL	DT	AJ	0011
NN	PP	VG	DT	0011	NN	PP	NN	C1	0011	NN	PP	DT	VP	0011	NN	NN	((	NN	0011
NN	NN	VL	DT	0011	DT	AJ	NN	,,	0011	C1	DT	AJ	NN	0011	AJ	NN	**	PN	0011
AJ	NN	NN	,,	0011	AJ	NN	NN	PP	0011	((	NN	))	,,	0010	,,	DT	NN	NN	0010
,,	C1	DT	AJ	0010	VP	TO	VV	DT	0010	VP	PP	NN	**	0010	VL	VP	P4	DT	0010
VL	DT	NN	PP	0010	VG	DT	NN	PP	0010	**	PN	NN	VL	0010	**	DT	NN	VL	0010
PP	PN	NN	**	0010	PP	NN	**	PN	0010	PP	AJ	NN	**	0010	NN	,,	C1	NN	0010
NN	,,	C1	AD	0010	NN	VL	AD	VP	0010	NN	TO	VV	DT	0010	NN	**	NN	PP	0010
NN	PP	NN	VP	0010	NN	NN	,,	VG	0010	NN	C1	NN	**	0010	DT	NN	AJ	NN	0010
DT	AJ	--	NN	0010	AJ	NN	VP	PP	0010	((	NN	))	VV	0009	((	NN	))	NN	0009
,,	PP	DT	NN	0009	VV	DT	NN	NN	0009	VG	DT	NN	NN	0009	PN	NN	VL	VP	0009
P4	NN	NN	,,	0009	NN	VP	PP	NN	0009	NN	**	PP	NN	0009	NN	RP	VV	DT	0009
NN	PP	PN	AJ	0009	NN	PP	NN	VG	0009	NN	NN	VP	PP	0009	NN	NN	**	AD	0009
NN	NN	P4	NN	0009	DT	VP	NN	NN	0009	DT	NN	NN	VV	0009	DT	AJ	NN	VV	0009
DT	AJ	NN	VP	0009	DT	AJ	NN	VL	0009	C1	NN	NN	**	0009	AJ	NN	NN	VL	0009
AJ	AJ	NN	,,	0009	AD	,,	DT	NN	0009	,,	((	NN	))	0008	,,	NN	,,	C1	0008
VV	TO	DT	NN	0008	VP	PP	NN	PP	0008	VL	VP	**	NN	0008	VL	AD	AJ	TO	0008

CONNECTION TABLE OF PARTS OF SPEECH 4--SYMBOL

				FREQ.					FREQ.					FREQ.					FREQ.
VG	DT	AJ	NN	0008	TO	DT	AJ	NN	0008	**	HH	VL	DT	0008	PP	NN	VL	VP	0008
PP	DT	NN	VP	0008	PP	DT	NN	AJ	0008	P4	DT	NN	NN	0008	NN	VL	VP	,,	0008
NN	**	HH	VL	0008	NN	PP	VP	NN	0008	NN	PP	RP	DT	0008	NN	NN	**	P4	0008
NN	NN	PP	PN	0008	NN	NN	--	NN	0008	NN	NN	AJ	NN	0008	--	NN	PP	NN	0008
--	NN	NN	,,	0008	--	NN	NN	PP	0008	HH	VL	DT	NN	0008	DT	NN	**	DT	0008
DT	NN	NN	P4	0008	DT	NN	--	NN	0008	DT	AD	AJ	NN	0008	C1	NN	NN	NN	0008
C1	NN	--	NN	0008	AS	VP	PP	NN	0008	AJ	NN	TO	VV	0008	((	NN	))	DT	0007
VH	PL	VP	PP	0007	TO	VV	NN	PP	0007	TO	VV	AJ	NN	0007	**	((	NN	))	0007
**	PP	PN	NN	0007	**	AD	,,	DT	0007	RP	VV	DT	NN	0007	PP	RP	DT	NN	0007
PP	NN	((	NN	0007	PP	NN	PP	AJ	0007	PP	NN	NN	C1	0007	PP	NN	C1	NN	0007
PP	DT	NN	TO	0007	PP	AJ	NN	,,	0007	PN	VL	VP	TH	0007	NN	VV	NN	PP	0007
NN	VG	DT	NN	0007	NN	**	PN	VV	0007	NN	NN	NN	((	0007	NN	))	NN	((	0007
NN	C1	NN	--	0007	DT	NN	,,	C1	0007	DT	NN	PP	RP	0007	DT	NN	NN	RP	0007
AJ	TO	DT	NN	0007	AJ	NN	**	NN	0007	AJ	NN	PP	VG	0007	AJ	NN	C1	NN	0007
AJ	AJ	NN	NN	0007	AD	PP	DT	AJ	0007	AD	DT	NN	PP	0007	AD	DT	NN	NN	0007
,,	PP	AJ	NN	0006	,,	NN	NN	NN	0006	VV	PP	AJ	NN	0006	VV	AD	PP	DT	0006
VP	**	NN	PP	0006	VL	VP	**	DT	0006	VL	VP	P4	NN	0006	TO	VV	DT	AJ	0006
TO	VL	VP	PP	0006	TO	DT	NN	NN	0006	**	NN	PP	NN	0006	**	DT	NN	VV	0006
PP	PN	AJ	NN	0006	PP	NN	,,	C1	0006	PP	NN	TO	VV	0006	PP	DT	VG	NN	0006
PP	DT	RA	NN	0006	PP	DT	NN	VL	0006	PP	AJ	--	NN	0006	PP	AJ	AJ	NN	0006
PN	VL	AD	AJ	0006	P4	DT	AJ	NN	0006	NN	,,	C1	VV	0006	NN	VL	AD	AJ	0006
NN	VG	PP	DT	0006	NN	TO	VV	NN	0006	NN	TO	VL	VP	0006	NN	TO	NN	NN	0006
NN	**	((	NN	0006	NN	**	PP	PN	0006	NN	**	P4	DT	0006	NN	PP	VG	NN	0006
NN	PP	NN	((	0006	NN	NN	VV	PP	0006	NN	NN	VH	PL	0006	NN	NN	TO	NN	0006
NN	NN	NN	VV	0006	NN	NN	NN	VH	0006	NN	--	NN	C1	0006	NN	))	**	DT	0006